

(美) D.弗雷费尔德 著

# 分子生物学

原核生物和真核生物总论

下 册

科学出版社





58.17  
165

# 分子生物学

原核生物和真核生物总论

下册

[美] D. 弗雷费尔德 著

俞梅敏 刘序倬 译

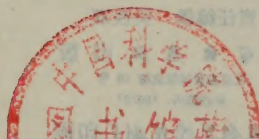
主编 王中

副主编 王中

校 不

中国科学院图书馆

中国科学院图书馆



中科院植物所图书馆



S0011794

科学出版社

1991

25752

## 内 容 简 介

本书共有三大部分：分子生物学研究中的生物系统和大分子的基本性质；大分子的生物合成和调节；真核生物系统的分子生物学。重点是原核生物和真核生物的基本分子过程（DNA、RNA 和蛋白质的生物合成）、遗传现象以及基因表达的调节控制。全书分上、下两册出版。下册内容包括噬菌体及质粒的分子生物学，真核病毒、肿瘤病毒及致癌基因的分子生物学，遗传物质的交换与重组——同源重组、转座作用及遗传工程，以及原核生物和真核生物中基因表达的调节控制。可供分子生物学、生物化学、遗传学、基础医学工作者以及高等院校有关专业师生参考。

David Freifelder

### MOLECULAR BIOLOGY

*A Comprehensive Introduction to Prokaryotes and Eukaryotes*

Science Books International, Van Nostrand Reinhold Company 1983

## 分 子 生 物 学

原核生物和真核生物总论

### 下 册

【美】D. 弗雷费尔德 著

俞梅敏 刘序倬 译

责任编辑 吴铁双

科学出版社出版

北京东黄城根北街 16 号

邮政编码：100707

中国科学院印刷厂印刷

新华书店北京发行所发行 各地新华书店经售

\*

1991 年 6 月 第 一 版

开本：787×1092 1/16

1991 年 6 月 第一次印刷

印张：21

印数：0001—1 900

字数：477 000

ISBN 7-03-002128-2/Q·294

定价：20.90 元



## 译 者 的 话

分子生物学是近 30 年来迅速发展形成的一门比较年轻的学科。它和其他许多学科(如生物化学、遗传学、微生物学、病毒学、细胞生物学、有机化学、物理化学、医学乃至技术科学)都有交叉。我国随着分子生物学领域研究工作的深入,高科技生物技术和生物工程研究的开展,以及高等学校相继设置分子生物学课程,广大研究人员和师生都迫切地需要参考书。为适应当前的需要,我们翻译了本书,它是目前国外在分子生物学方面最新最全面的一本教科书。

当前,分子生物学发展迅猛,正值本书翻译期间,作者 D·弗雷费尔德又推出了本书的第二版(1987)。为此我们又参照新版本对某些章节作了修改与增补。因新版本内容增多,所以将第一版(1983)中的真核病毒及肿瘤病毒(第二十一章)分别列为两章译出;而且还按新版本的体例,在肿瘤病毒一章中加入了致癌基因部分,使这一章作为比较新颖而完整的肿瘤分子生物学基础提供给读者。新版本真核生物基因及基因产物活性调节一章与第一版相比变化较大。除了保留第一版中已有的  $\phi$ X174 噬菌体、 $\lambda$  噬菌体、腺病毒、脊髓灰质炎病毒、逆转录病毒、SV40 病毒、病毒的转化作用、真核基因扩增、RecA 蛋白、同源重组作用机理、转座作用、真核细胞细胞质多聚蛋白质以及遗传工程等以外,还增添了不少新内容,例如蛋白质和核酸的相互作用机理(cAMP-CRP 和 cII 蛋白等与 DNA 的结合),基因拼接和 RNA 拼接对基因表达的调节,酵母的转化作用和基因表达的调节,甲基化及其在基因表达调节中的作用等,这些新内容均分别在下册中予以译出。译文中不妥之处,请读者批评指正。

译者

1989 年 12 月于北京





# 目 录

<b>第十四章 原核生物基因及基因产物活性的调节</b> .....	1
调节作用概述.....	1
基因活性调节的意义 .....	1
细菌代谢调节的一般机理 .....	2
转录调节的类型 .....	2
乳糖操纵子模型和乳糖体系.....	3
乳糖降解需要两个蛋白质的遗传学和生物化学证据 .....	3
lac体系受调节的证据: 诱导型和组成型的合成和阻遏 .....	4
o <sup>c</sup> 突变体的显性: 操纵基因 .....	5
操纵子模型 .....	6
本底水平的组成型合成 .....	7
Lac <sup>+</sup> 细菌培养物对乳糖的反应 .....	7
lac 阻遏物的纯化 .....	8
葡萄糖对乳糖操纵子活性的影响 .....	9
环式 AMP 与降解物激活蛋白 .....	10
与 cAMP 有关的其他体系 .....	12
乳糖操纵子 DNA 的调节区域 .....	12
cAMP-CAP 复合物及阻遏物的作用机理 .....	15
不完整体系的生理性质.....	15
从一个 lac mRNA 分子翻译出的 $\beta$ -半乳糖苷酶、透性酶及转乙酰基酶在数量上的差异.....	16
无诱导物存在时阻遏状态的解除 .....	16
乳糖操纵子与其他操纵子的融合 .....	17
半乳糖操纵子.....	18
gal 操纵子的遗传组成 .....	18
两个 gal 启动子以及 cAMP-CAP 对它们活性的影响 .....	19
cAMP-CAP 和 gal 阻遏物的作用机理.....	20
gal 操纵子中两个启动子的意义.....	20
阿拉伯糖操纵子.....	21
AraC 蛋白的正负调节作用 .....	22
AraC 蛋白的两种形式 .....	23
AraC 蛋白合成的调节 .....	23
ara 操纵子中调节结合位点的定位 .....	24
细胞营养状态对 ara 操纵子的影响 .....	25
色氨酸操纵子, 一个生物合成体系 .....	26
trp 操纵子的组成 .....	27
trp 的阻遏蛋白-操纵基因体系 .....	27
衰减子和前导多肽 .....	28



mRNA 前导序列中的碱基配对 .....	29
衰减的机理 .....	30
阻遏作用与衰减作用的对抗 .....	31
利用组氨酸的操纵子 .....	31
recA 操纵子 .....	33
核糖体合成的调节 .....	34
沙门氏杆菌鞭毛的相变异 .....	36
负反馈抑制 .....	38
基因产物对翻译的影响 .....	40
<b>第十五章 噬菌体 (I)</b> .....	42
烈性噬菌体 .....	42
噬菌体的一般性质 .....	42
噬菌体的结构 .....	42
典型噬菌体的裂解途径 .....	44
噬菌体感染的细菌培养物的性质 .....	46
噬菌体的必需和非必需基因 .....	48
噬菌体感染的专一性: 吸附作用 .....	49
噬菌体感染的专一性: 寄主限制 .....	50
噬菌体专论 .....	51
大肠杆菌 T4 噬菌体 .....	51
T4 DNA 的性质 .....	52
T4 噬菌体的基因构造 .....	54
T4 生活周期概述 .....	55
细菌大分子合成的关闭 .....	56
T4 噬菌体 DNA 的转录 .....	58
T4 DNA 的复制 .....	59
T4 噬菌体颗粒的产生 .....	62
大肠杆菌 T7 噬菌体 .....	65
T7 的基因构造 .....	66
T7 的生活周期 .....	67
T7 DNA 注入宿主细胞 .....	68
T7 DNA 转录作用的调控 .....	68
T7 DNA 复制 .....	70
T7 噬菌体的成熟 .....	70
两种枯草杆菌噬菌体 .....	71
噬菌体 PBS-2: 不需要宿主的聚合酶 .....	71
噬菌体 SPO1 .....	72
大肠杆菌 $\lambda$ 噬菌体 .....	72
$\lambda$ 噬菌体 DNA 及其注入后的转变 .....	72
$\lambda$ 基因的构造和调节位点 .....	73
裂解周期的时间顺序 .....	73
维持宿主体系完整性的必要条件 .....	74



$\lambda$ 的转录顺序 .....	74
晚期转录的时间顺序 .....	77
裂解周期中 $\lambda$ 转录的小结 .....	77
DNA 复制和成熟: 偶联过程 .....	78
大肠杆菌噬菌体 $\phi$ X174 .....	80
重叠基因 .....	82
吸附和穿透 .....	82
$\phi$ X174 DNA 的复制 .....	83
$\phi$ X174 DNA 的转录 .....	84
$\phi$ X174 DNA 的包装 .....	84
丝状 DNA 噬菌体 .....	85
单链 RNA 噬菌体 .....	87
一种双链 RNA 噬菌体: $\phi$ 6 .....	89
<b>第十六章 噬菌体(II)</b> .....	93
温和噬菌体及溶源生活周期 .....	93
溶源周期 .....	93
溶源生活周期概述 .....	93
一些有用的术语 .....	94
溶源及溶源化作用的一般性质 .....	94
大肠杆菌 $\lambda$ 噬菌体 .....	95
对感染的免疫性 .....	95
前噬菌体的整合 .....	97
诱导及前噬菌体的切除 .....	101
合成 $\lambda$ cl 阻遏物的两种启动子 .....	103
操纵基因的结构以及阻遏物的结合和 Cro 产物 .....	104
裂解周期还是溶源周期的决定 .....	106
溶源化作用过程中的切除及其防止 .....	107
用 $\lambda$ 筛选致癌物 .....	107
其他形式的溶源菌 .....	108
溶源菌的某些性质 .....	108
Mu——一种不寻常的噬菌体 .....	108
Mu DNA .....	109
Mu DNA 的复制与成熟 .....	110
转导噬菌体 .....	110
通用转导颗粒的性质 .....	111
通用转导颗粒的用途 .....	113
特化转导颗粒的性质 .....	114
从 $\lambda$ 溶源菌形成特化转导颗粒 .....	114
$\lambda$ gal 和 $\lambda$ bio 颗粒的性质 .....	115
非溶源菌的特化转导作用 .....	116
溶源菌的特化转导作用 .....	117
高频转导裂解物 .....	117



<b>第十七章 遗传物质重排与交换机理 (I)</b>	120
<b>质粒</b>	120
质粒的遗传学性质及类型	120
质粒 DNA: 性质及分离方法	121
质粒 DNA 的转移	124
转移的大致步骤及某些定义	124
F 质粒转移的细节	125
通过质粒 DNA 的转化作用	128
质粒的复制	128
对宿主蛋白质利用的差异	128
复制机理的差别	128
拷贝数的控制	129
不相容性	131
复制的抑制	132
特殊细菌质粒的性质	133
性质粒 (F 质粒)	133
抗药性质粒 (R 质粒)	136
产大肠杆菌素的质粒或 Col 质粒	137
其他细菌质粒	139
真核生物中的质粒	139
遗传工程中质粒的应用	140
<b>第十八章 遗传物质重排和交换机理 (II)</b>	142
<b>同源 DNA 序列间的重组</b>	142
微生物中重组的类型	142
同源重组的基本特征	144
同源重组的特殊性质	145
断裂-重接及杂合现象	145
分支移动	147
单链侵入	147
相互和非相互交换	148
单个重组事件子代的检测	149
<b>DNA 片段与染色体之间的重组</b>	150
细菌转化中的重组	150
细菌接合中的重组	151
转导作用中的重组	152
反式切割	153
<b>完整 DNA 分子之间的重组</b>	154
Holliday 结构	154
链转移或 Aviemore 模型	156
一个“先配对,后切割”的重组模型	159
$\delta$ 型分子的解体	160



RecA 和 RecBC 蛋白的性质 .....	161
RecA 蛋白 .....	161
RecBC 蛋白 .....	164
缺乏重组蛋白体系的性质 .....	164
大肠杆菌的其他重组体系 .....	165
<b>第十九章 遗传物质重排与交换机理 (III)</b> .....	168
可转位因子 .....	168
几个术语 .....	168
IS 序列 .....	169
IS 序列的发现 .....	169
IS DNA 的大小及性质 .....	169
IS 序列在大肠杆菌 DNA 和一些噬菌体 DNA 中的位置 .....	170
转座子 .....	171
转座子的发现 .....	171
转座子插入位点的定位 .....	173
转座子的类型 .....	174
转座作用 .....	176
插入位点的靶序列的复制 .....	176
转座子的结构和转座子-靶序列复合体的性质 .....	178
转座机理 .....	179
共整合 .....	179
在两个质粒之间发生转座时转座子的复制过程 .....	180
在 Tn3 的转座作用中共整合物作为中间物存在 .....	180
转座作用的两种模型 .....	181
转座子介导的其他遗传学现象 .....	184
复制子融合 .....	184
由转座子引起的缺失与倒转 .....	184
利用同一转座子两个拷贝之间同源性的过程 .....	186
在 Hfr 形成过程中 IS 序列的作用 .....	187
细菌中的基因扩增 .....	187
真核细胞中的转座子 .....	188
<b>第二十章 重组 DNA 及遗传工程</b> .....	190
DNA 分子的连接 .....	190
载体 .....	190
限制性内切酶 .....	191
测定切割位点;限制性内切酶图谱 .....	194
由许多限制性内切酶产生的具有互补单链末端的片段 .....	196
具有粘性末端的限制性内切酶片段的连接 .....	196
通过外加均聚物连接 DNA 片段 .....	198
平整末端的连接 .....	199
用连接器连接 .....	199
利用流体动力的剪切力产生 DNA 片段 .....	200

特定 DNA 分子加到载体中的插入作用 .....	200
cDNA 及逆转录酶的使用 .....	200
插入互补的人工合成基因 .....	201
用限制性内切酶切割大供体分子获得特定 DNA 片段 .....	202
重组 DNA 分子的检测 .....	202
保证质粒载体含有插入 DNA 片段的方法 .....	203
保证噬菌体载体含有外来 DNA 片段的方法 .....	206
鉴定含有外来 DNA 片段的质粒的物理学方法 .....	207
对于含有特定 DNA 片段的质粒的集群筛选 .....	208
基因文库 .....	211
从克隆的基因生产蛋白质 .....	212
生产真核生物蛋白质中的问题 .....	212
一个克隆基因的表达 .....	213
生长激素释放抑制因子的合成 .....	215
胰岛素原和胰岛素的合成 .....	215
从克隆的基因合成其他的真核生物蛋白质 .....	217
重组 DNA 技术的应用 .....	217
关于重组 DNA 的辩论 .....	219
<b>第二十一章 真核生物病毒</b> .....	221
真核生物病毒的基本结构 .....	221
病毒核酸 .....	224
病毒的 DNA .....	224
病毒的 RNA .....	225
烈性病毒的基本生活周期 .....	227
病毒吸附寄主细胞 .....	227
病毒核酸进入宿主细胞 .....	227
子代病毒颗粒的成熟和释放 .....	228
宿主细胞被病毒接管 .....	229
病毒颗粒的鉴定 .....	230
病毒的生物学鉴定 .....	230
检测病毒的效率 .....	231
动物病毒的分类 .....	232
动物 RNA 病毒 .....	234
病毒的 RNA 聚合酶 .....	234
泡状口炎病毒 .....	235
节片状(一)链病毒 .....	236
脊髓灰质炎病毒 .....	236
披膜病毒 .....	240
呼肠病毒 .....	241
逆转录病毒 .....	242
动物 DNA 病毒 .....	249
微小病毒 .....	249



乳多孔病毒:多瘤病毒和 SV40.....	251
大型 DNA 病毒.....	257
腺病毒 .....	257
疱疹病毒 .....	262
牛痘病毒 .....	263
关联病毒.....	263
感染性 DNA .....	264
干扰素.....	264
植物病毒.....	266
<b>第二十二章 肿瘤病毒和癌基因</b> .....	269
转化细胞.....	269
检测整合的病毒 DNA .....	271
整合的病毒 DNA 的结构.....	273
整合的可能后果.....	274
病毒的致癌基因.....	274
蛋白质激酶和逆转录病毒引起的转化 .....	274
癌细胞中的致癌基因 .....	275
src 在细胞中的对应体 .....	276
致癌物和致癌基因 .....	277
动物细胞中的致癌基因和病毒的致癌基因之间的关系 .....	278
致癌基因和致癌的多重步骤.....	279
致癌基因的活化.....	280
致癌基因蛋白质.....	281
蛋白激酶 .....	282
生长因子 .....	282
ras 蛋白 .....	283
DNA病毒中的转化蛋白质 .....	283
<b>第二十三章 真核生物基因及基因产物活性的调节</b> .....	286
原核细胞和真核细胞基因结构的重要差异.....	286
基因家族.....	287
简单的多基因家族 .....	288
复杂的多基因家族 .....	289
受发育控制的复杂多基因家族 .....	289
真核 mRNA 分子的多种类型 .....	292
真核基因表达的调节策略.....	292
基因改变.....	293
基因丢失 .....	293
基因扩增 .....	294
基因重排:免疫系统中编码序列之间的连接 .....	297
初级转录产物合成的调控.....	302
真核基因构造影响调控的特征 .....	302

酵母中的调节单位 .....	303
激素对转录的调控 .....	305
由转录因子介导的调控 .....	307
增强子活性的调节 .....	309
甲基化作用 .....	310
互补的转录单位 .....	311
加工过程的调节.....	312
多聚蛋白质: 同一 mRNA 产生多种蛋白质.....	314
翻译过程的调控.....	315
延长 mRNA 的寿命 .....	315
控制翻译的起始 .....	316
自我调节 .....	317
蛋白质合成总速度的调节 .....	317
光对植物基因表达的调节.....	319
卵黄蛋白原合成调节的一个例子.....	319
卵黄蛋白的合成 .....	319
卵清白蛋白的协同合成 .....	322



## 第十四章 原核生物基因及 基因产物活性的调节

不同的基因在单位时间内所编码合成的蛋白质分子的量不同。通过前面几章的学习,我们可以看到这是因为 RNA 聚合酶对启动子的识别效率以及起始翻译过程的效率不同而造成的。但是,这仅仅是遗传信息流(基因表达)调节的一个方面。例如,对于其他许多基因来说,生物体通常只是偶尔地需要这些基因的产物。因此,除非外界环境要求它们存在,否则在通常的情况下这些基因产物并不大量地存在。可见不仅各种基因产物在数量上存在着差异,而且在不同情况下,基因活性也不同,它是随时可以调节的。

在本章我们将讨论基因活性调节的意义以及代谢调控的基本原理,还将介绍一些在原核生物中已经了解得较为深入的调控体系。

### 调节作用概述

在这一节里我们将对调节体系的一般性质进行讨论。

#### 基因活性调节的意义

自然选择倾向于保留高效率的生命过程。在单细胞生物中,使细胞代谢总效率增加的突变会使突变型细胞的生长略快于野生型机体。这样,若有足够的时间,突变型细胞系的生长速度将超过野生型细胞系。例如,在一个生长速度为每 30 分钟增殖一倍的  $10^9$  细菌群体中,若有一个细菌的生长速度变成了 29.5 分钟增殖一倍,大约经过 80 天的连续生长后,这个群体中的 99.9% 都将具有 29.5 分钟增殖一倍的生长速度(在这个计算中,假设培养液被不断地稀释;除非培养液体积比地球还大,否则细菌将在一两天内停止生长)。这段时间,从实验室的角度来看是很长的,但相对于进化的时间尺度来说就微乎其微了。因此,从上述的分析来看认为调节体系的效率在进化过程中得到一定的提高,是有道理的。

一个细胞有两种提高代谢效率的方式:(1) 形成一个新的体系,该体系所需要的自由能较低,或者简单地提高反应速度;(2) 尽量减少浪费。第二种方式是本章所要叙述的重点。

细胞内的调节作用具有以下几个或多或少带普遍性的规律:

1. 某些偶尔才需要的分子,仅仅当这种需要增长时才合成出来。
2. 如果一个酶的作用只是在无意义地消耗能量,或者消耗一些可以作为另一个酶反应底物的物质,那末,该酶的活性通常要受到抑制。
3. 如果一个细胞在利用能量时有几种降解途径可供选择,必将选择在单位时间内可产生较大能量的那种途径。

4. 改变一个生物合成途径,使之减少生成有缺陷的分子的机会,以提高效率,因而这种调节是有利的。

细菌代谢调节的一般机理

细菌中最基本的调节机理遵循以下规律：一个体系的活性在需要时被开启，而在不需要时被关闭。这种“开启-关闭”型的活性调控是通过调节基因的转录作用，即允许 mRNA 合成或阻止 mRNA 合成，或调节转录酶的活性来实现的。事实上，人们并没有见到过活性完全关闭的体系。当一个转录过程处于“关闭”状态时，其基因表达还总是保留在基础的水平上。在这种情况下，每代细胞常常大约进行一两次转录，因此这时蛋白质的合成量极低。为方便起见，我们在讨论转录时将使用“关闭”这一词，但这实际意味着极低的转录水平。在这一章里我们还会看到一些体系的例子，它们的活性可从全部地开启转变到部分地开启(或甚至是微弱地开启)，而不是完全关闭。

在细菌中，如果在某一个代谢途径中有若干个酶依次地起作用，这些酶往往或者全部出现，或者全不出现。这种现象称为协同调节。这种协同现象是对一个多顺反子 mRNA 的合成过程进行调节而引起的，这个 mRNA 编码所有这些酶。形成这种调节作用的几种机理将在下一节介绍。

酶活性一般由反应产物的浓度调节，或者在生物合成过程中，由一系列反应的最终产物调节。这种调节称为反馈抑制或终端产物抑制。小分子的效应物也常常用来激活或抑制某种特定酶的活性。酶活性的调节将在本章更后的部分来讨论。

转录调节的类型

转录水平上的调节有几种常见的方式，这取决于某一体系代谢活性调节的类型。例如，在降解代谢途径中，起始端的酶(一般指催化第一步反应的酶)的底物浓度往往决定是否合成这一途径中的其他各个酶。相反，在合成代谢中，最终产物则往往是调节物质。甚至在一个从单顺反子 mRNA 翻译出蛋白质的体系中，该蛋白质就可以充当自我调节的角色，即启动子的转录活性直接由该蛋白质的浓度决定。每一调节类型的分子机理相差甚远，但一般可以分为负调节和正调节两大类。

在负调节体系中，细胞中存在着抑制物，它使转录处于关闭状态。抗抑制物通常称为诱导物，它在转录的开启中发挥作用。在正调节体系中，一个效应物分子(它可以是蛋白质、小分子化合物或几个分子的复合物)激活启动子，而抑制物在正调节中也并未失去其作用。正负调节之间并不相互排斥，有些系统中既有正调节也有负调节，因而需要两种调节物。表14-1列出了正负调节的特点。

表 14-1 调节的类型

调节物结合到 DNA 上	正 调 节	负 调 节
是	开 启	关 闭
否	关 闭	开 启

注意：如果某一体系兼有正负调节，当正调节物结合到 DNA 上时，正调节开启，负调节物就不结合到 DNA 上。



降解体系可以具有正负两种调节。而生物合成过程往往由最终产物对自身的合成进行负调节,最简单的方式是产物缺少就促进合成,而产物过多则抑制合成。

真核生物中的调节规律不同于细菌中的方式,例如真核细胞中所有的 mRNA 都是单顺反子。研究真核细胞的调节机理很困难,但随着重组 DNA 技术的发展,情况已有很大改善。这方面内容将在后面的章节再行介绍。

在本章中我们要讨论细菌中各种各样的调节体系,将从研究得最透彻的体系开始,即先介绍与乳糖利用有关的调节,因为乳糖体系的研究提供了理解调节机理所需要的术语和重要原则。

## 乳糖操纵子模型和乳糖体系

对代谢调节的详细研究始于乳糖降解体系,大多数描述调节机理的术语来自 Francois Jacob 与 Jacques Monod 对该体系的出色的遗传学分析。我们的介绍从分析他们所观察到的现象开始。

### 乳糖降解需要两个蛋白质的遗传学和生物化学证据

我们现在知道在大肠杆菌中有两种蛋白质是乳糖利用过程所必需的,一个是  $\beta$ -半乳糖苷酶,其作用是将乳糖裂解,生成半乳糖和葡萄糖(见第二章);另一个是乳糖透性酶,它是乳糖分子进入细胞所必需的。遗传学实验与生物化学分析相结合,首次证明这两种蛋白质确实存在。

首先分离出几百个  $\text{Lac}^-$  的突变体(不能以乳糖作为碳源而生存)细菌,通过遗传重组的方法,把一些突变部分从大肠杆菌的染色体上转移到  $\text{F}'\text{lac}$  的性质粒上(这种性质粒携带着可以利用乳糖的  $\text{lac}$  基因),然后构建出具有  $\text{F}'\text{lac}^-/\text{lac}^+$  的或  $\text{F}'\text{lac}^+/\text{lac}^-$  的局部二倍体的基因型质粒(质粒的基因型写在斜线左边,染色体的基因型则写在右边)。这些二倍体都产生  $\text{Lac}^+$  的表现型(即它们都能够合成  $\beta$ -半乳糖苷酶),说明  $\text{lac}^-$  突变体没有产生抑制物来阻碍  $\text{lac}$  基因的活性。后来,又构建了一些局部二倍体,它们的染色体与  $\text{F}'\text{lac}$  质粒都是  $\text{lac}^-$  的突变体。对所有以上不同组合配对的  $\text{lac}^-$  突变体进行研究发现,某些细菌表现型为  $\text{Lac}^+$ ,而另一些则为  $\text{Lac}^-$ 。这种互补实验(见第一章)表明最初分离到的突变体可归纳为分别称为  $\text{lacZ}$  和  $\text{lacY}$  的两类(为简单起见,在讨论乳糖体系时,我们将使用一个字母来代表遗传因子,这样,基因  $\text{lacA}$ ,  $\text{lacY}$ ,  $\text{lacZ}$  以及非编码区“位点”  $\text{lacO}$ ,  $\text{lacP}$  和  $\text{lacI}$  就被简写为  $a, y, z, o, P, i$ )。属于  $y$  和  $z$  的突变体经过互补而构成局部二倍体时,如果是  $\text{F}'y^-z^+/y^+z^-$  或  $\text{F}'y^+z^-/y^-z^+$  的形式,就有  $\text{Lac}^+$  的表现型;但如果是  $\text{F}'y^-z^+/y^-z^+$  和  $\text{F}'y^+z^-/y^+z^-$ , 则表现为  $\text{Lac}^-$ 。以上两种互补类型的存在有力地证明了在乳糖体系中至少有两个基因在调节作用中发挥作用。

将细胞暴露于  $^{14}\text{C}$  标记的乳糖中,放射性物质不能进入  $y^-$  细胞,但很容易被  $z^-$  细胞所吸收。用溶菌酶来处理  $y^-$  细胞。这种酶可以部分地破坏细胞壁使之透性大为增加,于是处理后的  $y^-$  细胞可以吸收  $^{14}\text{C}$  乳糖。其他一些实验证明  $z$  基因是  $\beta$ -半乳糖苷酶的结构基因,例如对  $z^-$  突变体的提取液进行免疫学分析,发现存在着一种与  $\beta$ -半乳糖苷酶结构相似但不完全相同的蛋白质,而且没有酶活性。通过对遗传学图谱进行分析,

进一步得到  $\gamma$  和  $z$  基因是相邻的证据。

**lac 体系受调节的证据：诱导型和组成型的合成和阻遏**

乳糖代谢体系的开启-关闭性质是通过以下观察得到的：

1. 如果具有  $lac^+$  基因型的大肠杆菌生长在不含乳糖及  $\beta$ -半乳糖苷的培养基中，细胞内的  $\beta$ -半乳糖苷酶和透性酶的浓度则将很低，每个细胞内大约只有两个酶分子。但是，如果在培养基中加入乳糖，酶的浓度则将达到细胞总蛋白质的百分之一（每个细胞中约有  $10^5$  个酶分子）。

2. 如果把乳糖供给无乳糖（也无葡萄糖，关于这点很快将讨论到）的培养基中生长的  $Lac^+$  细菌培养物， $\beta$ -半乳糖苷酶和透性酶将同时合成，如图 14-1 所示。进一步用  $^{32}P$  标记的 mRNA 作实验，使从杂交噬菌体  $\lambda lac$ （即带有  $lac$  基因的  $\lambda$  噬菌体）中取得的 DNA 与在乳糖加入后的不同时间内所形成的  $^{32}P$ -mRNA 进行分子杂交，结果表明加入乳糖能激发  $lac$  的 mRNA 的合成。

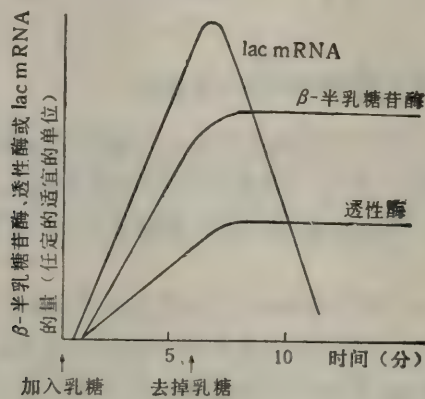
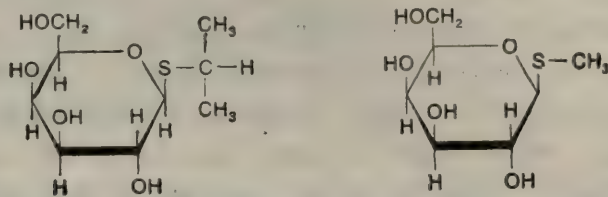


图14-1  $lac$  体系的“开启-关闭”性质。在加入乳糖之后，很快就显现出  $lac$  mRNA；在该 mRNA 合成滞后一些时候，显现同时产生出  $\beta$ -半乳糖苷酶和透性酶，因为翻译是需要时间的。当移去乳糖之后，不再形成  $lac$  mRNA，而已有的  $lac$  mRNA 的量却因 mRNA 的正常降解而减少。 $\beta$ -半乳糖苷酶和透性酶都是稳定的，甚至在不再有以上两种酶合成时，它们的浓度仍然维持恒定。

以上事实说明乳糖体系是可诱导的，而且乳糖是一个诱导物。科学家对于此诱导作用机理的研究奠定了我们对代谢调节认识的基础。

在研究诱导作用时很少使用乳糖，因为乳糖会被诱导合成的  $\beta$ -半乳糖苷酶所催化裂解，因而乳糖浓度在实验中不断下降，使许多问题复杂化（如动力学实验）。作为替代物，常常使用两种含硫的乳糖类似物：



异丙基硫基半乳糖苷 (IPTG)                      硫甲基半乳糖苷 (TMG)



这两种物质是有效的诱导物,却不是  $\beta$ -半乳糖苷酶的底物。具有这种性质的诱导物称为安慰诱导物 (gratuitous inducers)。

后来又分离到一种在有诱导物或没有诱导物的情况下都能产生 lac mRNA 的突变体,这种失去调节能力的突变体称为组成型 (constitutive) 突变体。Jacob 和 Monod 用这种组成型突变体构建了各种局部二倍体的细胞,并将这些突变株归纳为两类: i 和  $o^c$ 。特点见表 14-2。

表 14-2 带有 i 和  $o$  突变型的不同组合的局部二倍体的性质

基 因 型	组 成 型 (C)	可 诱 导 型 (I)
1. $F' o^c z^+ / o^+ z^+$	C	—
2. $F' o^+ z^+ / o^c z^+$	C	—
3. $F' i^- z^+ / i^+ z^+$	—	I
4. $F' i^+ z^+ / i^- z^+$	—	I
5. $F' o^c z^+ / i^- z^+$	C	—
6. $F' o^c z^- / o^+ z^+$	—	I
7. $F' o^c z^+ / o^+ z^-$	C	—

$i^-$  突变株的行为像大多数基因的突变(—)那样,并且是隐性的(表 14-2 的 3、4)。当  $i^+$  细胞停止 mRNA 的合成时,  $i^-$  突变体仍然允许 mRNA 进行合成。因此很显然, i 基因是一个产生抑制物的调节基因,其产物使体系关闭。  $i^-$  突变体由于不能产生这种抑制物,因此成为组成型的。在  $i^+ / i^-$  的局部二倍体中有一个正常的 i 基因产物拷贝,所以此体系仍然可以被抑制。Jacob 和 Monod 称 i 基因产物为 lac 阻遏物 (lac repressor)。最初人们见到这种情况还不清楚 lac 阻遏物究竟是一种蛋白质还是一种 RNA 分子(指的是由 i 基因转录形成的,但不能编码蛋白质的 RNA)。后来发现此基因有琥珀突变体(使蛋白质链的合成停止的突变),有力地证明了阻遏物是一种蛋白质。1967 年,人们首次分离到了阻遏物并进一步使之纯化,其性质以后再讨论。遗传学图谱分析指出, i 基因紧接着 z 基因,从而确定了基因 i、z、y 的顺序。

**$o^c$  突变体的显性: 操纵基因**

$o^c$  突变体的一个显著特征是它们在某些情况下表现出显性(表 14-2 的 1、2、5)。这些  $o^c$  突变体的显性特征的重要意义在某些局部二倍体中更加清楚(表 14-2 的 6、7)。这两种组合体由于有功能正常的 z 基因而表现为  $lac^+$ 。但在 6 中我们能看到,尽管有  $o^c$  突变存在,  $\beta$ -半乳糖苷酶的合成仍然是可诱导型的。两个组合的差异在于 6 中的  $o^c$  突变与  $z^-$  突变处于同一个 DNA 分子上,而 7 中在同一个 DNA 分子上却有着  $o^c$  与  $z^+$ ; 所以只有当  $o^c$  与  $z^+$  在同一 DNA 上时,  $o^c$  才能引起  $\beta$ -半乳糖苷酶的组成型合成。这一结论的确证主要来自一个重要的生物化学实验: 检测  $\beta$ -半乳糖苷酶突变型的免疫学实验中,表明在  $o^c z^- / o^+ z^-$  的局部二倍体中(表 14-2 的 6)有突变酶被组成型地合成出来,而只有在加入诱导物之后,才会合成出野生型的酶\*。关于  $o^c$  突变还有一个现象: 遗

\* 实验是这样进行的: 首先制备纯化的  $\beta$ -半乳糖苷酶的抗体。野生型的这种抗体和突变体在结构上的差异不是太大,它也可以和突变的  $\beta$ -半乳糖苷酶进行反应。这种反应称之为交叉反应 (cross-reaction), 突变体蛋白质被称为交叉反应物质 (CRM)。CRM 可以通过一系列标准的免疫学步骤检测出来,因此,它可以作为突变体蛋白质是否存在的指示物。

遗传图谱分析指出  $o^c$  突变位于  $i$  与  $z$  之间,所以  $lac$  体系的四个基因的次序为  $iozya$ 。通过这些观察, Jacob 和 Monod 推断  $o^c$  突变代表 DNA 链上的一个位点或一个非编码区域,而不是一个基因,因为可编码的基因具有互补性。 $o$  决定相邻的  $z$  基因的产物是诱导型合成还是组成型合成。 $o$  区域称为操纵基因。

### 操纵子模型

Jacob 和 Monod 提出用操纵子模型来解释  $lac$  体系的调节机理(见图 14-2),这个模型现在已被人们广泛接受。

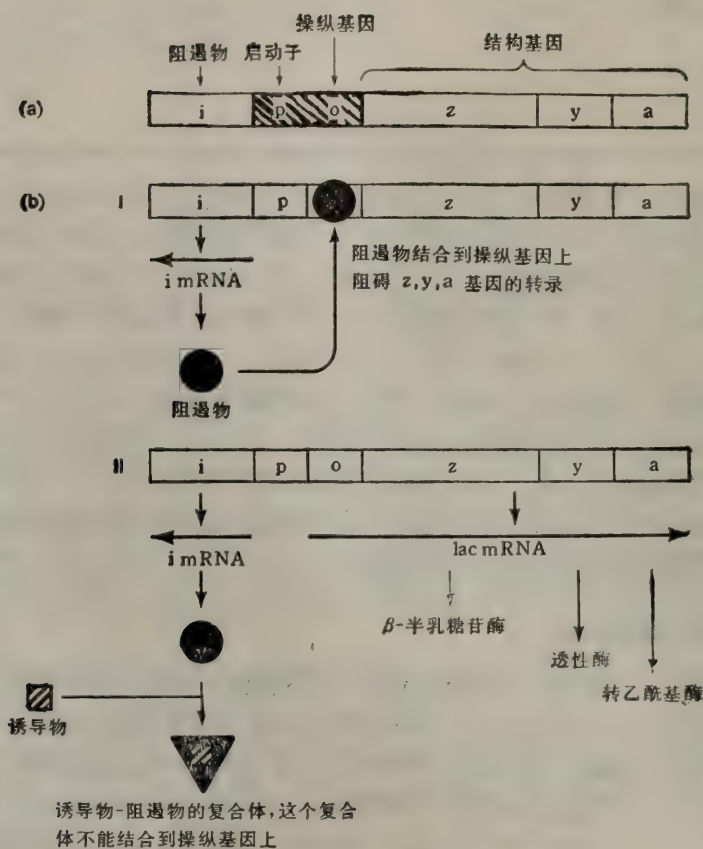


图 14-2  $lac$  操纵子。(a) 遗传图谱。请注意未按尺度绘制, 事实上的  $P$  位点和  $o$  位点要比其他基因小得多。(b) 图示: I. 阻遏状态; II. 诱导状态。诱导物改变阻遏物的形状, 使阻遏物不再能与操纵基因结合。

1.  $z$ 、 $y$  基因的产物是由同一条多顺反子的 mRNA 分子编码的\*。

\* 这个 mRNA 分子还包含有第三个基因的转录产物,该基因以  $a$  表示,  $a$  编码转乙酰基酶,该酶催化乳糖分子中的半乳糖部分的乙酰化作用。 $a^-$  突变体可以正常进行乳糖代谢,是  $lac^+$ 。转乙酰基酶在乳糖代谢中不起作用,在此暂且忽略,其功能以后再讨论。



2. 这个 mRNA 分子的启动子紧接着  $\sigma$  区域。几年之后分离到突变型的启动子( $p^-$ ), 并确定它位于  $i$  与  $\sigma$  之间, 它完全不能启动合成  $\beta$ -半乳糖苷酶和透性酶。

3. 操纵基因是 DNA 上的一段序列, 是阻遏物结合的位点。

4. 当阻遏物与操纵基因结合时, lac mRNA 的转录起始受到抑制。

5. 诱导子通过与阻遏物结合而激发 mRNA 的合成\*, 这种结合改变了阻遏物的三维构象, 使之不能与操纵基因结合, 这样, 有诱导物存在时, 操纵基因区域没有被占领, 所以启动子能够使 mRNA 合成起始, 并进行下去。

这一简单模型解释了 lac 体系以及其他负调节体系的许多性质。但在后一节里我们将会发觉这种解释仍是不完善的, 因为乳糖操纵子中同样也存在着正调节。

### 本底水平的组成型合成

有两个矛盾是操纵子理论所不能解释的。首先, 诱导物需要穿过细胞膜去与阻遏物结合, 而转运诱导物又需要透性酶, 后者的合成又需要诱导。这样我们必须解释诱导物第一次是如何达到细胞内的。只有两种可能, 或者一些诱导物可以在透性酶不存在时进入细胞, 或者一些透性酶可以在没有诱导物的情况下合成, 我们很快可以看到第二种解释是正确的。

另一个矛盾是我们近年来发现乳糖(葡萄糖-1,4-半乳糖)并不和阻遏物结合, 真正的诱导物是乳糖的异构体——异构乳糖(葡萄糖-1,6-半乳糖), 但是异构乳糖是在  $\beta$ -半乳糖苷酶的催化下由乳糖形成的, 因此, 乳糖诱导  $\beta$ -半乳糖苷酶的合成需要有  $\beta$ -半乳糖苷酶的存在。这两个表面上反常的现象有着同样的解释: 在非诱导状态下有少量的 lac mRNA 合成(大约每代有一个 mRNA 分子合成出来), 这种合成称之为本底水平的组成型合成(background constitutive synthesis), 由于阻遏物的结合也不是绝对紧密的, 即使是在它与操纵基因紧密地结合时, 偶尔会掉下来; 这时启动子的障碍被解除, mRNA 聚合酶开始转录。这种现象以每代细胞一次的概率发生。

### Lac<sup>+</sup> 细菌培养物对乳糖的反应

现在我们可以用分子水平的术语来描述当少量乳糖加到生长着的 Lac<sup>+</sup> 培养物中所发生的现象。假使细菌生长在以甘油为碳源的培养基中, 每个细菌中有 1—2 个分子的  $\beta$ -半乳糖苷酶和乳糖透性酶。这时加入乳糖, 在单个的透性酶的作用下, 少量乳糖分子进入细胞, 又在单个的  $\beta$ -半乳糖苷酶分子的作用下, 乳糖转变成异构乳糖。某个异构乳糖与操纵基因上的阻遏物结合, 并使后者失活而离开基因, 这样 mRNA 的合成就开始了。这些 mRNA 分子又合成大量的  $\beta$ -半乳糖苷酶及乳糖透性酶, 其结果是乳糖涌入细胞。多数乳糖被降解并产生葡萄糖和半乳糖; 但还有许多乳糖被转变成异构乳糖, 然后与细胞内的阻遏物结合(虽然合成速度低, 但是阻遏物仍继续地在合成着, 因此通常是有足够的异构乳糖使细胞维持着去阻遏状态的), 阻遏物的失活使 mRNA 的合成以高速度进行, 透性酶和  $\beta$ -半乳糖苷酶的浓度也非常高。裂解产生的葡萄糖被细胞用作碳源和能源; 裂解产生的半乳糖则被一套酶体系转变成葡萄糖, 这些酶的合成也是可诱导性的。这

---

\* 由于有了这个现象, 我们在很多体系中使用去阻遏来描述诱导。

个可诱导的体系被称为 *gal* 操纵子,半乳糖是它的诱导物,这些将在本章的另节进行讨论。

当培养基和细胞中的所有乳糖都被消耗完了的时候,由于阻遏物还在不断地合成着,有活性的阻遏物的浓度将超过异构乳糖的浓度,细胞重新建立起阻遏状态, mRNA 合成被抑制。在细菌中,多数 mRNA 的半寿期只有几分钟(见第十三章);在不到一个世代的生长时间内,细胞内已几乎没有 *lac* mRNA,  $\beta$ -半乳糖苷酶及透性酶的合成也趋向停止,这些蛋白质虽然很稳定,但其浓度随着细胞的分裂而不断被稀释。值得注意的是,如果在原有的乳糖被撤去之后的一个世代中再加入乳糖,这时乳糖可以立即开始裂解,因为此时细胞内仍有一定浓度的透性酶和  $\beta$ -半乳糖苷酶。

### *lac* 阻遏物的纯化

证明操纵子模型的主要假说的一个重要步骤是分离 *lac* 阻遏物并研究其性质。这一工作是利用对放射性 IPTG (一种安慰诱导物) 的结合来标记阻遏物而完成的。这种结合通过平衡透析法进行检测,见图 14-3。从细胞中提取出经过部分纯化的、不同组分的蛋白质(即含有阻遏物等),让它们与 IPTG 结合,发现结合能力不同。*lac* 阻遏蛋白具有四个相同的亚基,各个亚基均含有 347 个氨基酸残基,每个亚基可与一分子 IPTG 结合。粗的、未经分级分离的细胞提取物的结合能力为大约每个细胞结合 20—40 个 IPTG 分子,因此大约每个细胞中有 5—10 个阻遏物分子。进一步观察到在某些 *i<sup>-</sup>* 突变株的细胞提取物中缺少阻遏物,从而证明了与 IPTG 结合的分子确实是阻遏物。

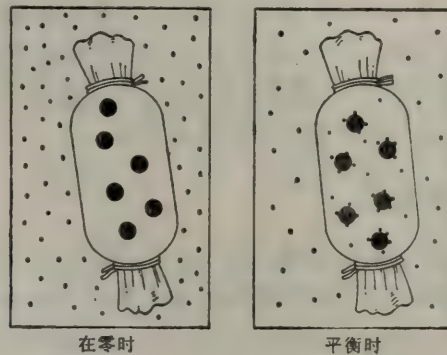


图 14-3 平衡透析法。将装着大分子物质(图中的大颗粒)的透析袋放到含有 IPTG (图中的小颗粒)的溶液中, IPTG 可与阻遏物结合。平衡时,透析袋内外的游离 IPTG 浓度相等。因为阻遏物能结合某些小分子物质,因此小分子物质(此为 IPTG)在透析袋内的浓度大于它在袋外的浓度。

由于阻遏物的数量极少,以至认为每个世代中的这些蛋白质分子是由每个细胞内不超过 1—2 个被转录的阻遏物的 mRNA 模板翻译出来的。mRNA 的数量是如此之少,因此很可能阻遏物合成受到调节,也可能这种 mRNA 是从一个弱启动子转录出来的。在其他一些操纵子的阻遏物调节中已看到过这两种机理在发生作用,但是对乳糖操纵子来说只有后者才是正确的,即阻遏物的 mRNA 在一个弱启动子的控制下组成型地合成出来。关于形成阻遏物分子的数量很少的原因,可以用几个突变体的性质清楚地说明,因为



在这些突变体中弱启动子转变成了强启动子。细胞内不可能有足够的诱导物来克服阻遏状态,因此这些突变体是不可诱导的。这种过量产生阻遏物的细菌对于实验研究是非常宝贵的,因为能产生高浓度的阻遏物(占细胞总蛋白量的1%)就意味着可以大量地纯化它们,使人们可以对该种蛋白质进行物理性质的研究,并对它的氨基酸序列进行分析。

利用纯化的阻遏物,人们已经证明它与操纵基因序列的专一性结合,以及诱导物对于这种结合的抑制。这些研究中所用到的 lac DNA 来自一个噬菌体的变种(转导噬菌体,见第十六章),它是经过遗传操作后携带有大肠杆菌的 lac DNA 的噬菌体。图 14-4 示出一个纯 lac 阻遏物结合到 DNA 分子上的电镜照片。

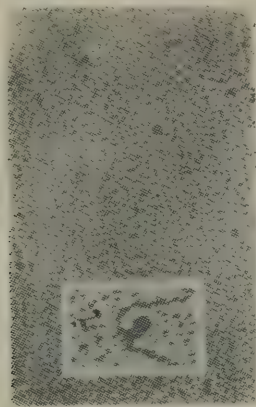


图 14-4 大肠杆菌 lac 阻遏物的电子显微镜照片(放大 107 000 倍)。图中显示了一个阻遏物分子结合到 lac DNA 上的景象(放大 215 000 倍)(承 Robley Williams 提供)。

研究阻遏物-操纵基因结合的一个重要方法是硝酸纤维素滤膜分析,蛋白质能够吸附到滤膜上,而DNA不能。当阻遏物与<sup>14</sup>C-DNA的混合物滤过这种膜时,如果这时有DNA-蛋白质的复合物形成,复合物中的<sup>14</sup>C就能够停留在膜上。表 14-3 是这个实验的结果。这些结果指出阻遏物结合到操纵基因上(即阻遏物不结合到突变的操纵基因 o<sup>c</sup>上),IPTG 妨碍这种结合。以上这些结果证实了操纵子模型的主要假说部分。

表 14-3 阻遏物与操纵基因的结合

通过滤膜的结合物	<sup>14</sup> C 结 合 于 膜
<sup>14</sup> C-lac DNA	否
<sup>14</sup> C-lac DNA + 阻遏物	是
<sup>14</sup> C-lac DNA + 阻遏物 + IPTG	否
<sup>14</sup> C-lac DNA(o <sup>c</sup> ) + 阻遏物	否

葡萄糖对乳糖操纵子活性的影响

β-半乳糖苷酶在乳糖代谢中的作用是降解乳糖,形成葡萄糖及半乳糖(如前面已提到的,同时半乳糖将在 gal 操纵子的作用下再转化成葡萄糖)。如果将葡萄糖和乳糖同时加

在培养基中，从效益的观点出发，对于细胞来说没有必要诱导乳糖操纵子。事实确实如此，细胞就是按照这种逻辑表现它的行为，即在外源葡萄糖耗尽之前，不形成  $\beta$ -半乳糖苷酶（见图 14-5）。当有葡萄糖存在时，不合成  $\beta$ -半乳糖苷酶，其原因是没有 lac mRNA 的形成。对 mRNA 转录的抑制可能来自两个方面，(1)葡萄糖阻止了诱导物引起的阻遏物失活的作用，因此在这种情况下阻遏物有活性；(2)不只是移走阻遏物，还必须有其他因素的参与才能起始 mRNA 的合成。我们将会看到第 2 个假设是正确的。

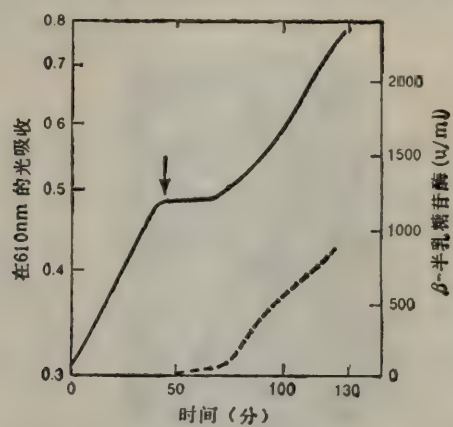


图 14-5 葡萄糖代谢转变为乳糖代谢的实验培养基中同时含有葡萄糖及乳糖的细菌培养物，在葡萄糖耗尽时，糖代谢转变到乳糖代谢方面。图中箭头所指处葡萄糖不再有供应，细胞数量不再增加。因为当有葡萄糖存在时细菌不合成 lac mRNA，所以也不产生  $\beta$ -半乳糖苷酶。当葡萄糖浓度下降时，cAMP 浓度增加，直到开始合成  $\beta$ -半乳糖苷酶(下面一条曲线)。610nm 的光吸收用于测量细胞的总量；酶活性用任定的单位表示。

葡萄糖对乳糖操纵子表达的抑制作用是间接的，关于这种作用可以由一个大肠杆菌的突变株的性质证明。这个突变株在酵解途径中不能将葡萄糖-6-磷酸转变为下一步的代谢中间物。这种突变株可以在葡萄糖存在时被诱导（例如被 IPTG）合成出 lac mRNA。因此，很显然不是葡萄糖而是葡萄糖的某些降解产物抑制了 lac mRNA 的合成。由于这样的实验结果，因此葡萄糖的效应又被称之为降解产物的阻遏效应（catobolite repression）（我们还将看到，对于一个并不利用阻遏物进行调控的过程来说，这是一个使用不当的名称）。在前一节介绍的例子中，我们讨论了将乳糖加入到以甘油为碳源的基本培养基（minimal medium）中时，乳糖操纵子的诱导效应。甘油只能在葡萄糖-6-磷酸以后的阶段参加到酵解代谢途径中，所以我们所关心的降解产物是在甘油进入到该步骤之前就已存在。现在还没有得到关于实际降解产物的进一步资料。然而，正如我们就将看到的，葡萄糖效应的基本原理已经弄清楚了。

### 环式 AMP 与降解物激活蛋白

1957 年 Sutherland 发现环式 AMP(cAMP)，并发现它在动物组织中是广泛存在的（图 14-6）。



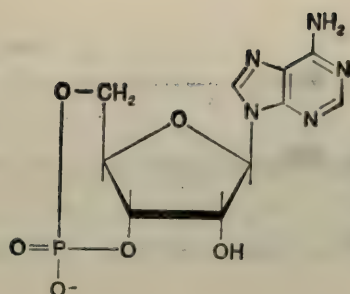


图 14-6 cAMP 的结构。

cAMP 是在腺苷酸环化酶作用下由 ATP 转变而来的。在多细胞的真核生物中，cAMP 在许多激素的作用过程中起着重要作用。它也存在於大肠杆菌中，其浓度受葡萄糖代谢的调节。如果将细菌放在缺乏能源的培养液中进行培养（例如放在无碳源的培养液中保温培养），细胞内 cAMP 的浓度非常高。如果培养在含有葡萄糖的培养液中，则 cAMP 浓度就很低。但是如果培养液中只含有甘油或其他不能进入酵解途径的碳源物质，cAMP 的浓度也会很高（表14-4）。对这一现象可能的解释是，某种酵解途径中位于葡萄糖-6-磷酸与甘油进入位点之间的代谢物质，或者其他途径中的代谢产物是腺苷酸环化酶的抑制物（inhibitor）。这些结果的重要意义在于指出了 cAMP 是乳糖操纵子活性的一个中介体（mediator），

表 14-4 在不同碳源的培养基上生长的细胞中 cAMP 的浓度

碳 源	cAMP 的 浓 度
葡萄糖	低
甘油	高
乳糖	高
乳糖+葡萄糖	低
乳糖+甘油	高

大肠杆菌（其他细菌中也有可能）含有一个称为代谢产物活化因子蛋白（catabolite activator protein, CAP）的蛋白质，它由 *crp* 基因编码，能够与 cAMP 形成一个复合体，在 *crp* 基因及腺苷酸环化酶基因上有突变的突变株细菌都不能合成 lac mRNA，因而使人们认识到 CAP 及 cAMP 都是 lac mRNA 合成所必需的。进一步的研究指出，CAP 能够与 cAMP 形成一个复合物，记作 cAMP-CAP，这是乳糖体系中一个有活性的复合物。细菌对于此复合物的需要是独立的，与阻遏体系无关，因为 *crp* 基因和环化酶基因突变的突变株细菌即使同时也是 *i*-或 *o*<sup>c</sup> 突变，仍然不能合成出 lac mRNA。我们现在知道进行转录必须有 cAMP-CAP 复合物结合在 DNA 的启动子区域上。因此，cAMP-CAP 是一个不同于阻遏物的正调控因子，而乳糖操纵子的功能则是在这两个相互独立的正、负调节体系的作用下实现的（图 14-7）。

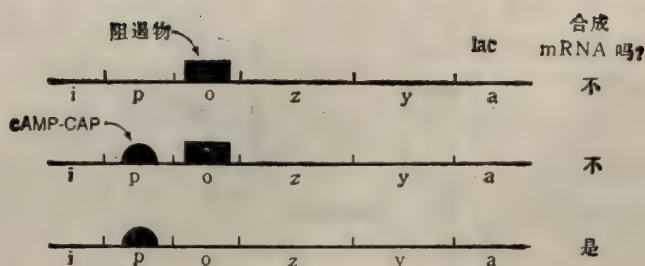


图 14-7 乳糖操纵子的三种状态示意图。表示仅仅在存在 cAMP-CAP 和缺少阻遏物时才形成 lac mRNA。

## 与 cAMP 有关的其他体系

还有许多其他的糖(如半乳糖、麦芽糖、阿拉伯糖、山梨醇等),它们在降解过程中被转化成葡萄糖或酵解途径中的其他中间产物。这些糖中的每一个糖的糖代谢中有关的酶都是由可诱导的操纵子合成的;可以想像得到,如果有葡萄糖存在,这些操纵子都不能被诱导,它们被称作降解产物敏感型操纵子(catabolite-sensitive operons)。有个简单的遗传学实验可以证明这些操纵子都是由 cAMP-CAP 调节的。在一个糖代谢的操纵子上,以  $10^{-6}$  的频率出现单一突变(如  $\text{lac}^-$  或  $\text{mal}^-$ ),双突变  $\text{lac}^- \text{mal}^-$  出现的频率则是  $10^{-12}$ ,这在实验中是无法检测出来的。但是表现型为  $\text{Lac}^- \text{Mal}^-$  或  $\text{Gal}^- \text{Ara}^-$  的双突变却能够以可以检测的频率出现。这种明显的双突变并非由两个操纵子上的突变引起,而常常是因为  $\text{crp}^-$  或不能合成腺苷酰环化酶所致。进一步发现,如果一个突变体是  $\text{Lac}^- \text{Mal}^-$ ,它同时常常也是  $\text{Gal}^- \text{Ara}^-$ 、 $\text{Lac}^- \text{Ara}^-$  等等。对少数几个降解产物敏感型操纵子进行生化分析,结果表明在这些体系的每一个操纵子中都有 CAMP-CAP 结合在启动子区域。

阿拉伯糖(以及某些其他五碳糖)的分解代谢中需要 cAMP 的原因还不很清楚,因为这些糖并不转化为葡萄糖。这说明 cAMP 的作用还有其更为一般的特征。我们还记得前面讲过,葡萄糖作为最理想的碳源,其浓度决定着细胞内的 cAMP 浓度。当葡萄糖的供应不很充分时,大量地产生 cAMP;即大量 cAMP 的出现是葡萄糖不足的一个信号。当葡萄糖耗尽时, cAMP 浓度升高使很多糖代谢的操纵子处于备用状态。因此,一旦这些糖中的某种糖提供给细胞,操纵子就能迅速地作出反应,如果葡萄糖又出现,那么信号就不存在了,因为这时细胞已没有必要再处于备用状态。

## 乳糖操纵子 DNA 的调节区域

利用前面研究核小体(见第六章)和 RNA 聚合酶结合位点(第十一章)时所用到的核酸酶-保护技术(nuclease-protection procedure),我们可以分离到含有乳糖操纵子的 DNA 片段。在这个技术的操作过程里,首先将纯化的阻遏物吸附到从转导噬菌体(称为  $\lambda\text{lac}$ )中提取出的 DNA 上(该种噬菌体携带着大肠杆菌的乳糖操纵子)。然后加入 DNase,将得到一段在酶解反应之后仍然保持完整的很短的 DNA 片段(24 碱基对)。这个片段的碱基序列以及用其他的方法得到的一个稍大的片段的碱基序列见图14-8。这种序列是



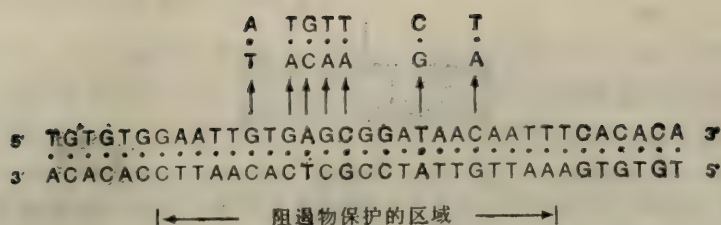


图 14-8 *lac* 操纵子的碱基序列。对称的有关区域用灰色表示。竖箭头指出所观察到的突变性改变,这些碱基改变使操纵子变为组成型的。

很特别的,人们常常会遇见这种序列。它具有被调节因子所识别的碱基序列的特点,即倒转重复顺序 (inverted repeat)\*。关于这个发夹结构就是操纵基因的证据来自于对操纵基因突变体的研究。因为阻遏物不能与突变的操纵基因结合,所以就无法用核酸酶-保护技术分离出与阻遏物相结合的突变的操纵基因。但是利用重组 DNA 技术 (见第二十章)可以分离得到含有操纵基因的较长的片段,根据图 14-8 中操纵基因所在位置的特点,就可以在这个片段中鉴定出操纵基因。这个方法同时也可用以确定操纵基因两翼的一些碱基顺序。这时如果在一些已经鉴定过的两翼序列之间找到一般与野生型序列有关的序列,就可以鉴定突变的操纵基因,它们可能只与野生型相差 1 或 2 个碱基。实验工作者从携带着不同  $o^c$  突变的几种  $\lambda$  *lac* 噬菌体中得到一些有关片段,并观察到如图 14-8 所示的碱基变化。注意,无论单个碱基变化发生在倒转重复序列或者发生在其他的碱基序列中,都会改变甚至消除阻遏物与操纵基因结合的能力。

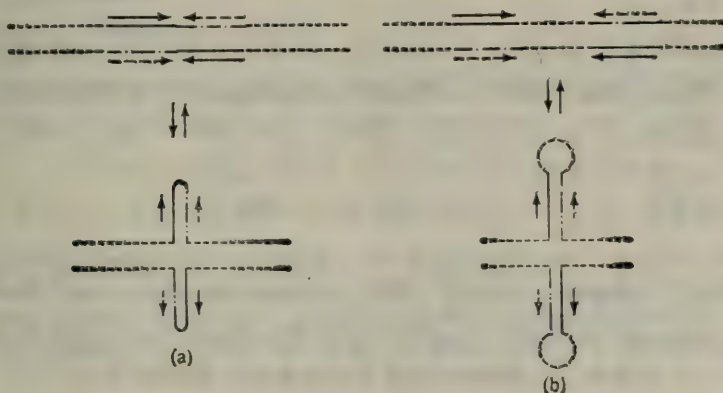
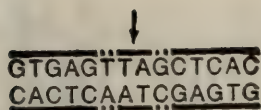


图 14-9 具有两个倒转重复序列的 DNA 分子可能具有的不同形式。(a) 两个倒转重复序列是紧紧相邻的; (b) 两个倒转重复序列由一个间隔区隔开。水平箭头表示序列的方向。

用核酸酶-保护技术还可以分离到 cAMP-CAP 结合位点, 这个位点的序列如下:

\* 这个名词表示一种像 ABCDEE'D'C'B'A 的序列, 在此序列中 A 及 A' 代表可以形成普通碱基对的碱基。这种序列又可称为回文结构 (palindrome)。人们设想, 并且已在某种情况下证实, 在体内 (in vivo) 这种序列可能可以偶然形成双发夹 (double hairpins) 或双柄与泡 (double-stem and loop) 的结构 (见图 14-9)。



注意,除去有点的两个碱基之外,还有一个倒转重复序列,分别对称地列于箭头两侧。通过遗传学及生理学实验已证明有两种“启动子向下坠落”(promoter-down)的突变(即这些突变阻止合成 mRNA)就是这种序列中的碱基发生了变化而引起的,这种突变阻碍了 cAMP-CAP 的结合,这同时也证实了 cAMP-CAP 的结合是转录作用从 lac 启动子处开始的前提条件。

图 14-10 示出分离出来的含有 cAMP-CAP 结合位点、RNA 聚合酶作用位点和操纵基因及其两翼序列的 DNA 的一个长片段。通过测定纯化的 lac mRNA 5' 端的碱基序列,从而在上述 DNA 长片段中找到与之互补的序列,这样便确定了 mRNA 合成的起始位点。 $\beta$ -半乳糖苷酶的起始密码子及 i 基因的终止序列,则从已知的  $\beta$ -半乳糖苷酶和阻遏物的氨基酸序列而得到确定。这个完整区域的特点有几个值得注意的性质:

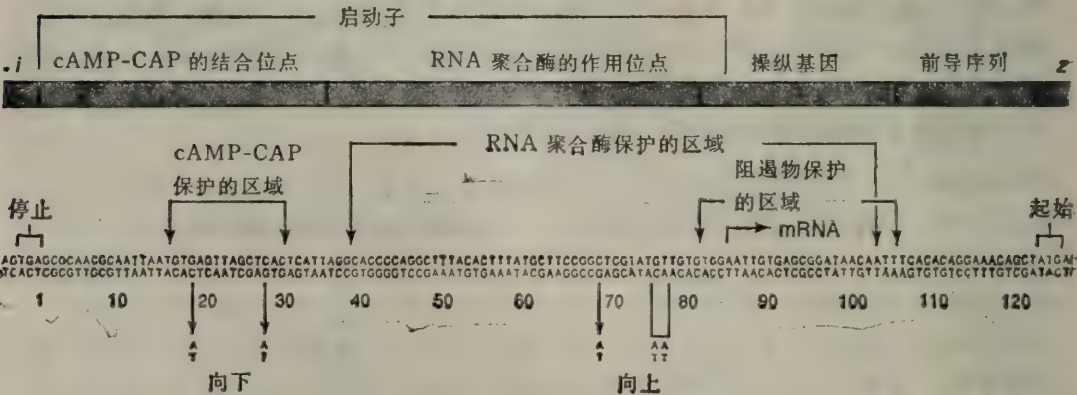


图 14-10 大肠杆菌 lac 操纵子的调节区的 DNA 碱基序列。下图中的阴影区表示 cAMP-CAP 结合位点中的倒转重复序列。图上表示了 DNA 上受各种蛋白质所保护的、能抵抗 DNA 酶消化的区域。图中也指出了几种碱基改变使启动子发生突变:产生一种无活性的启动子(向下坠落),或产生一种有特别高活性的启动子,或是产生一种不依赖于 cAMP-CAP 的启动子。图中还分别指出了在 z 基因中的起始密码子和 i 基因中的终止密码子。

1. mRNA 合成的起始位点在操纵基因中,当阻遏物蛋白存在时,起始位点被阻遏物占据。
2. 当 RNA 聚合酶结合在操纵基因上时,这一段结合序列可以免受 DNase 的降解。
3. cAMP-CAP 的结合位点非常接近于 RNA 聚合酶的作用位点,但不与之重合。
4. 调节区域的长度为 115 碱基对,约为 DNA 螺旋的 12 圈。
5. 这一区域的顺序为 ipoz, 这个顺序与遗传学分析的结果是一致的。



## cAMP-CAP 复合物及阻遏物的作用机理

lac mRNA 合成的起始必须有 cAMP-CAP 结合在启动子上,而且这时阻遏物必须与操纵基因结合(见图14-7)。目前对 cAMP-CAP 激活转录和阻遏物抑制转录的机理还没有详细地了解。但是,从对于 DNA 中的与 RNA 聚合酶和与阻遏物相接触的碱基序列(这些复杂的实验不准备在这里叙述)的分析,以及从以下的事实,已经推论出一个合理的模型:

1. 在体外, cAMP-CAP 的存在促进 RNA 聚合酶与 lac DNA 的结合。
2. 如果 RNA 聚合酶已结合到 lac DNA 上,那么接着再加入阻遏物,并不能把 RNA 聚合酶取代下来。

3. 如果阻遏物已先与操纵基因结合,那么 RNA 聚合酶便不再能形成稳定的 RNA 聚合酶-操纵基因复合体,但是在这个位点上还是会发生瞬间的微弱结合。

从上面给出的信息,可以对转录的起始机理作出以下假设:

1. 像第十一章中所介绍的那种复合物一样, RNA 聚合酶可以和 lac 启动子区域形成松散的复合体。

2. cAMP-CAP 的结合很可能使 DNA 双螺旋结构松弛,从而促进形成一个稳定的开放型启动子复合体(第十一章)\*。

3. 阻遏物是一个抗融解蛋白质(antimelting protein),这样它就能够阻止形成开放的复合体。也就是说,结合着的阻遏物可能阻碍 RNA 聚合酶去接近开放型启动子中的碱基。

4. 当开放型的启动子复合体形成时, RNA 聚合酶就启动 lac mRNA 的合成。

## 不完整个体的生理性质

如果调节机理对一个细胞来说是至关重要的,那么失去调节能力的突变型操纵子在某些方面就会有很多缺陷。到目前为止,在  $i^-$  及  $o^c$  突变体上所观察到的不足,只是生长速度略微降低。虽然这种降低只是在实验室里才能看到,但是从进化的角度上讲则是很意义的。一个生命体系如果由于具备了调节功能而提高了生长速度,尽管这种增加是很小的,可是对于细胞来说却是一种非常重要的优势——这些细胞能够因此而在自然选择中保留下来。

在某些条件下(可以想像这些现象在自然界中发生得会更为频繁),组成型 lac 突变体的生长速度显著地下降。对这些条件进行分析时,人们发现在 lac 操纵子中还应当有第四个基因,即我们前面曾提到过的  $a$  基因。该基因编码  $\beta$ -半乳糖苷转乙酰基酶,乳糖本身不能被乙酰化,但是这个酶能够把乙酰基从酰基供体上转移到许多个半乳糖苷分子上。我们将会看到这是一个非常重要的性质。

在自然界中存在着许多种能被  $\beta$ -半乳糖苷酶降解的、被取代的半乳糖苷分子。这些被取代的半乳糖苷的分解产物往往不能参加进一步的代谢,因而在体内积累到很高的浓度。这在很多情况下都是有害的,因为有好几种半乳糖苷衍生物在高浓度时是细胞正常生

---

\* 近来在物理学实验中已经观察到, DNA 的双螺旋的稳定性在离某个蛋白质结合位点有一定距离的一个位点上发生动摇,这种现象被称为远距稳定性 (telestability)。

长的抑制物,例如作为诱导物的 IPTG,这种半乳糖苷衍生物虽然并不是天然物质。人们发现,在 IPTG 存在时,当细菌是  $i^-$  和  $o^c$  突变体同时又是  $a^-$  的突变体时,它的生长显然要比相应的  $a^+$  对应体的生长缓慢得多, $a^-$  与  $a^+$  细胞的区别在于  $a^+$  细胞中 IPTG 是乙酰化的,而乙酰化的 IPTG 不能被  $\beta$ -半乳糖苷酶所分解,如下所示:

$i^+o^+$ :

$$\text{IPTG} \xrightarrow{\text{低}[\beta\text{-半乳糖苷酶}]} \text{低浓度的抑制性切割产物} \rightarrow \text{微弱的抑制作用。}$$

$i^-a^-$  或  $o^c a^-$ :

$$\text{IPTG} \xrightarrow{\text{高}[\beta\text{-半乳糖苷酶}]} \text{高浓度的抑制性切割产物} \rightarrow \text{抑制作用。}$$

$i^-a^+$  或  $o^c a^+$ :

$$\text{IPTG} \xrightarrow{\text{转乙酰基酶}} \text{乙酰化-IPTG} \xrightarrow{\beta\text{-gal}} \text{没有降解产物} \rightarrow \text{没有抑制作用}$$

因此  $a$  基因可能并不在乳糖的降解过程中起作用,但它能够在  $\beta$ -半乳糖苷酶存在时防止抑制性的半乳糖衍生物在细胞内的积累。

人们自然会想到的一个问题,为什么细胞在甚至只有乳糖作为唯一的半乳糖苷时,还要花费能量去合成可能被独立调控的转乙酰基酶?这个酶独立地受到调节(即不与  $\beta$ -半乳糖苷酶合成的调节相偶联)显然是不合理的,因为只有在  $\beta$ -半乳糖苷酶存在时才需要转乙酰基酶。也许乳糖操纵子中还有一个二级调节体系,但这可能需要花费更多的能量。可能在自然界中,当细菌生长在以乳糖作为基本碳源的环境中时被取代的半乳糖苷大量地存在着,因此这时  $a$  基因直接偶联到  $z$  基因上是十分必要的。

## 从一个 lac mRNA 分子翻译出的 $\beta$ -半乳糖苷酶、透性酶及转乙酰基酶在数量上的差异

在一个完全被诱导的细胞中, $\beta$ -半乳糖苷酶、透性酶及转乙酰基酶的拷贝数比例为 1:0.5:0.2。作为几百万年进化过程演变的结果,这个比例在一定程度上反映了在以  $\beta$ -半乳糖苷作为唯一碳源时细胞的需要。不同的酶在数量上的差异是由于在翻译水平上受到调节所致。这种调节有以下两种方式:

1. lac mRNA 经常离开它进行翻译的核糖体,从而终止蛋白质链的翻译过程。这种现象发生的频率取决于每一个后续 AUG 密码子再度起始翻译的概率。因此,就存在着一个从 mRNA 的 5' 末端到 3' 末端的蛋白质合成梯度。对于大多数多顺反子的 mRNA 分子发挥功能来说确实存在着这种现象。

2. lac mRNA 在内切酶作用下发生的降解,在  $a$  基因中比在  $z$  基因中更加频繁些。因此,在任何时候, $z$  基因的完整拷贝要比  $a$  基因的多。

在这一章里我们会看到,这种调节模式是经常发生的,即一个操纵子活性全面表达受到的调控作用是通过控制多顺反子 mRNA 的转录来实现的,而该 mRNA 所编码的各种蛋白质的相对浓度则由每一个顺反子翻译频率受到的调控作用所决定。

## 无诱导物存在时阻遏状态的解除

如果 lac 操纵子处在无阻遏物的细胞中,转录将会发生。我们可以通过将 lac 操纵子从基因型为  $i^+z^+$  的 Hfr 雄性细胞(见第一章)转移到  $i^-z^-$  的雌性细胞(这种细胞既



不形成阻遏物,也不产生  $\beta$ -半乳糖苷酶),当  $\text{lac DNA}$  进入没有阻遏物的雌性细胞时,细胞内立即产生  $\text{lac mRNA}$ ,从而在雌性细胞中也随之合成出  $\beta$ -半乳糖苷酶(见图 14-11)。又因为在这种雌性细胞里,转移进来的  $\text{lac DNA}$  中的  $i$  也被转录,产生  $i^+ \text{mRNA}$ ,因此也能够合成阻遏物(虽然相对于  $\beta$ -半乳糖苷酶来说合成速度要低得多)。因此,在这种情况下 $\beta$ -半乳糖苷酶的合成不能无限制地继续进行下去。现在已经在所有的负调控的操纵子中观察到这种现象。

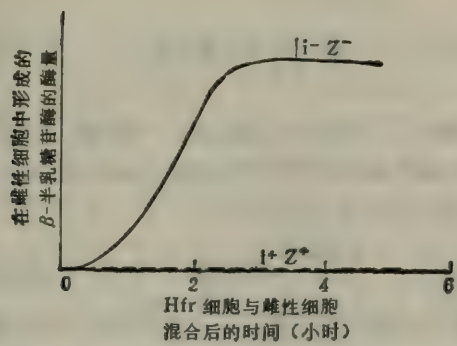


图 14-11  $\text{lac DNA}$  由  $i^+z^+$  的 Hfr 雄性细胞转移到  $i^-z^-$  或  $i^+z^-$  的雌性细胞后,在雌性细胞中合成  $\beta$ -半乳糖苷酶。在  $i^+z^-$  的细胞中不合成该酶,因为在这种细胞中总是存在着阻遏物。在  $i^-z^-$  的细胞中该酶的合成到一定程度终止,最终建立起阻遏作用。

### 乳糖操纵子与其他操纵子的融合

基因融合在自然界能不能发生,对研究来说是非常有价值的。一个研究得较多的例子是  $\text{lac}$  操纵子与  $\text{pur}$  操纵子的偶联,后者是负责嘌呤合成的基因。这个操纵子以完全不同于乳糖操纵子的方式进行调节,但我们在这里不打算讨论这种调节的机理。 $\text{pur}$  操纵子在大肠杆菌的染色体上位于  $\text{lac}$  操纵子的下游(沿转录方向)(见图 14-12)。在这两个操纵子之间的是  $\text{tsx}$  基因,这个基因控制着细胞对 T6 噬菌体的敏感性。从  $\text{tsx-s } z^+$  细胞中分离到  $\text{tsx-r}$ ,同时它又是  $z^-$  的突变体。这种突变体具有的缺失:从  $\text{lac}$  操纵子的  $z$  基因内开始,一直包括了所有  $\text{tsx}$  基因,并延伸到  $\text{pur}$  操纵子中,及至越过  $\text{pur}$  的操纵基因和启动子的一部分。这种缺失使  $\text{lac}$  操纵子失去了 RNA 聚合酶在  $\text{lac}$  操纵子中的终止序列,所以当转录作用从  $\text{lac}$  启动子处开始起始时,就合成出由一部分  $z$  基

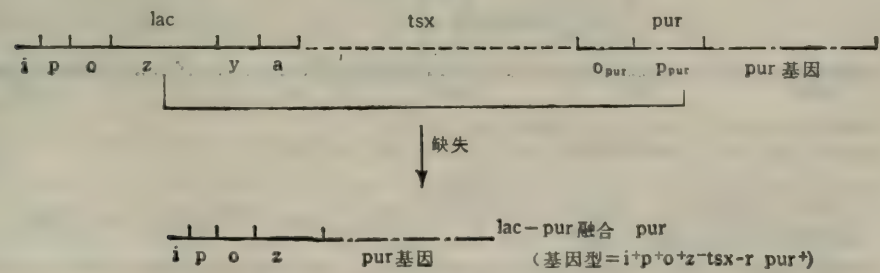


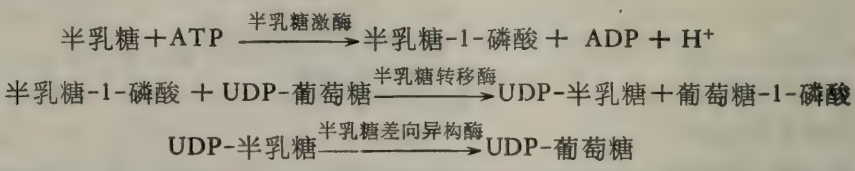
图 14-12  $\text{lac-pur}$  的融合及其起始点。

因区域再加上所有的 pur 基因所转录出的 mRNA 分子。因此,在这个融合体系中, pur 操纵子的转录过程可以受  $\beta$ -半乳糖苷的诱导。

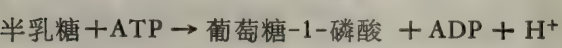
利用将 lac 操纵子在大肠杆菌染色体上移位 (relocation) 的技术可以将很多基因与 lac 启动子融合。这种技术对于生物化学家来说是非常有用的,因为 lac 启动子是一个很强的启动子,通过它可以使较弱启动子的 mRNA 的转录增强,从而增加蛋白质的合成量。在某些情况下,从细菌培养液中经过提取得到某种特定蛋白质的回收量可以增加上千倍。

## 半乳糖操纵子

半乳糖也是大肠杆菌可以用来作为碳源的另一种糖。它的代谢过程由三个酶完成:半乳糖激酶、半乳糖转移酶及半乳糖差向异构酶 (galactokinase, galactose transferase 及 galactose epimerase), 它们的作用顺序如下:



总的反应式为:



半乳糖操纵子 (gal operon) 的调节,原则上与 lac 操纵子的调节相似,但还存在着一些很有意思的细节差异,这些差异反映了半乳糖在细胞代谢过程中的双重作用。

### gal 操纵子的遗传组成

相应于乳糖代谢的三个酶 (I,II,III) 的基因分别为 galK, galT 和 galE。它们聚集在一起,共用一个操纵基因 (gal o)、一个启动子区域 (gal p) 以及一个阻遏物基因 (gal R), 以上这些基因共同组成了一个操纵子。

gal 操纵子与 lac 操纵子最显著的不同是 gal R (阻遏物的结构基因) 远离包括操纵基因、启动子和 gal 基因的基因簇(见图 14-13)。事实上因为阻遏物是一种可以扩散的

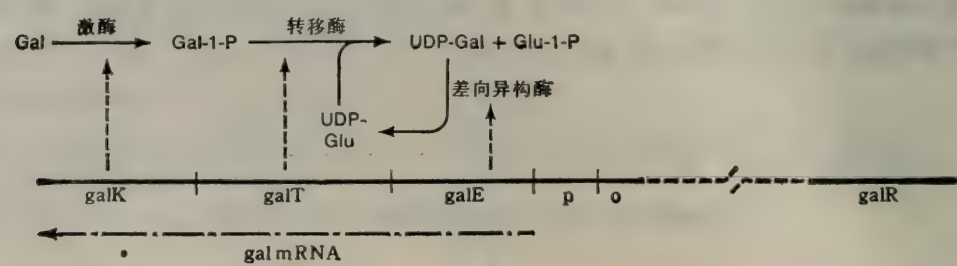


图 14-13 大肠杆菌 gal 操纵子的有关酶及因子。缩写: Gal, 半乳糖; Gal-1-P, 半乳糖-1-磷酸; UDP-Gal, 尿嘧啶二磷酸半乳糖; Glu-1-p, 葡萄糖-1-磷酸; UDP-Glu, 尿嘧啶二磷酸葡萄糖。gal R 阻遏物基因离操纵子的其他部分很远。垂直的直线箭头将基因与对应的酶联系起来。



物质，所以没有什么理由要求阻遏物基因必须与操纵基因紧密相邻。并且实际上也证明了这一点：自然界中阻遏物基因与结构基因直接相邻的操纵子与这类两种基因相互分开的操纵子几乎是一样多。*gal* 阻遏物(目前尚未纯化)的作用很像 *lac* 阻遏物的作用，那就是 *gal*  $R^-$  与 *gal* $^o$  突变体都具有组成型的表现型。当然，操纵基因与启动子(一个以上)与那些总是以适宜的方式排列着的结构基因相邻接。诱导物很可能不是某个代谢产物而是半乳糖本身。

两个 *gal* 启动子以及 cAMP-CAP 对它们活性的影响

半乳糖不能像葡萄糖那样有效地被利用，这使人们以为在葡萄糖存在时半乳糖操纵子会处于非诱导状态，然而，事情恰恰相反，即使有葡萄糖存在，这个操纵子也是可诱导的。关于这点的理由将很快就会明白。人们已经分离到若干个启动子的突变体，令人惊讶的是，这些突变体仅仅可归纳为两类，一类突变体只是在葡萄糖存在时才失去高水平合成 *gal* 酶的能力；另一类在葡萄糖不存在时也不能进行酶的合成。从 *gal* 操纵子具有两个启动子的事实出发，以上两种现象都能够得到解释。从这两个启动子开始合成的 mRNA 分子只是在碱基序列的长度上相差五个核苷酸，所以两个转录子的 DNA 中都含有 *galK* 基因、*galT* 基因和 *galE* 基因；事实上这两个 mRNA 所含的信息容量是同样的。进一步了解这两个启动子的性质需要分析启动子-操纵基因区域的碱基序列。

人们已经分离得到含有启动子-操纵基因区域的具有 139 个碱基长度的 DNA 片段，并测定了该片段中适当部分的碱基序列（见图 14-14）。图中表示出了两个 *gal* mRNA 分子的起始点 S1 和 S2，这些位点的确定是通过分离 mRNA 分子并测定其 5' 末端序列。S1 和 S2 在功能上有以下不同之处：

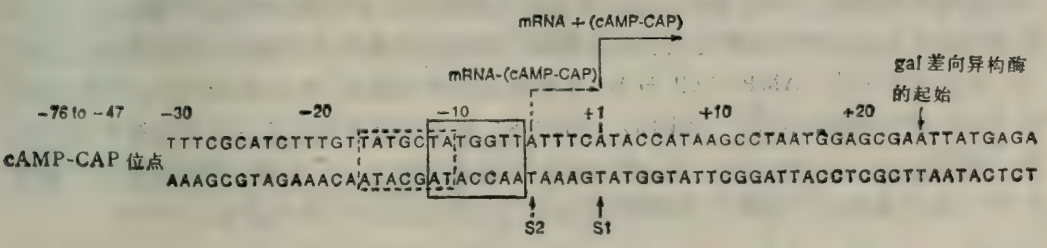


图 14-14 *lac* 操纵子和 *gal* 操纵子的 cAMP-CAP 结合位点。在这两个操纵子中具有相似碱基序列的区域用虚线标出。

1. 只在没有葡萄糖存在时才能从 S1 位点开始发生转录；在 *crp* $^-$  和腺苷酸环化酶的突变体中，都不能进行转录。在体外，只有在有 cAMP-CAP 存在时才能产生转录子。
  2. 有葡萄糖存在时，从 S2 起始的转录可以进行。这一过程不需要 cAMP-CAP 的存在；事实上，后者甚至对转录有抑制作用。
- 显然，葡萄糖不能对半乳糖操纵子的诱导作用进行抑制，因为这个操纵子中有一个起始位点并不需要 cAMP-CAP，即使在有葡萄糖存在的情况下，该位点也处于活跃状态。因而在有葡萄糖存在的情况下，阻断合成那些 *gal* 酶的启动子的突变必然是在 S2 中发生的突变；而那种在没有葡萄糖时不能进行酶合成的突变，就必定是 S1 位点中的突变。

**cAMP-CAP 和 gal 阻遏物的作用机理**

利用核酸酶-保护技术,人们已鉴定了 cAMP-CAP 作用的位点,该位点与 S1 的距离与乳糖操纵子中的相应距离大致相同,而且碱基序列也表现出一定程度的对称性,但不如在乳糖操纵子中这个区域的对称性那么严格。也许有人以为 cAMP-CAP 在所有操纵子上的结合位点都是相同的(因为 DNA 都是与同一种蛋白质结合),但是事实并非如此,在 lac 和 gal 操纵子中该序列有一定的相似性(见图 14-15),但在后面将要讨论到的阿拉伯糖操纵子中,它的结合序列就与乳糖操纵子大不一样了。cAMP-CAP 激活 S1 而抑制 S2 的机理尚不清楚。有一种观点认为 S1 受激活的方式与乳糖操纵子相似(即促进形成一个开放型的启动子复合物),而对在 S1 位点起始的转录过程的激活可能会干扰从 S2 起始的转录。

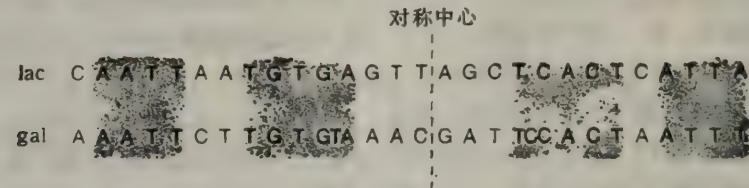


图 14-15 lac 操纵子与 gal 操纵子上 cAMP-CAP 的结合位点。图中阴影部分表示两个操纵子中的上述区域具有类似的碱基序列。

gal 阻遏物的作用机理也是非常复杂的,所有已知的操纵基因突变都发生在 cAMP-CAP 的结合部位点中,而这个结合位点对于 S1 和 S2 的两个 Pribnow boxes (即 RNA 聚合酶的结合位点,见第十一章)则相距甚远。在这种情况下 gal 操纵子的阻遏物不可能像乳糖操纵子中的 lac 阻遏物那样去阻碍 RNA 聚合酶的结合。现在认为 gal 阻遏物结合在 cAMP-CAP 位点上,这种结合可能以两种方式阻断从 S1 的转录起始作用:或者通过干扰 cAMP-CAP 与 DNA 的相互作用来阻断,或者通过改变 cAMP-CAP 对 RNA 聚合酶结合的效应而阻断。而在 S2 位点的阻遏作用似乎有着全然不同的机理。如果有 gal 阻遏物存在,那么转录的起始作用能够在 S2 发生,但当转录向下游进行 10—20 个碱基之后便停止了,这个过程是怎么发生的,原因尚不清楚。

**gal 操纵子中两个启动子的意义**

为什么半乳糖操纵子需要两个转录起始位点?对这个问题还没有明确的答案,但这可能与半乳糖在细胞代谢中的双重作用有关。半乳糖不仅可以作为唯一碳源供细胞生长,而且与之相关的物质尿苷二磷酸半乳糖 (UDPGal) 是大肠杆菌细胞壁合成的前体。在没有外源半乳糖的情况下,UDPGal 是通过半乳糖差向异构酶 (galactose epimerase) 的作用由 UDP-葡萄糖合成的,该酶是 galE 基因的产物。因此,如果细胞是生长的,那么在生长过程中的所有时间里细胞必须能够合成差向异构酶,这个作用也许可以通过不属于 gal 操纵子的第二个 galE 基因来完成。另一方面,因为在合成细胞壁的过程中对异构酶的需要量很小,本底水平的组成型合成可能就已经可以满足需要了。实际上 gal mRNA 的组成型合成水平已高于乳糖操纵子所合成的 lac mRNA 的水平,显然这个



mRNA 是在没有半乳糖的情况下作为酶合成的模板而存在的。

设想如果只有 S1 一个启动子,那么由于这个启动子的活性依赖于 cAMP-CAP,当葡萄糖出现时异构酶就不能组成型地产生。假设如果唯一的启动子是 S2,这种情况下即使在有葡萄糖存在的情况下半乳糖也将使操纵子处于充分脱阻遏(诱导)状态,这对细胞来说无疑是一种浪费。所以从必要性及经济上来考虑,都需要一个不依赖于 cAMP-CAP 的启动子(S2)进行本底水平的组成型合成,以及一个依赖于 cAMP-CAP 的启动子(S1)对高水平合成进行调节;进一步说,只有在 S2 活性完全被 cAMP-CAP 抑制时调控作用才是有效的\*。

对于 gal 酶合成的调节机理的研究还很不充分,这里还涉及到另一个调节因子,已通过在一个叫做 capR 的基因中发生的突变对此调节因子做过鉴定(注意 capR 基因与 CAP 蛋白毫无关系)。capR<sup>-</sup> 突变体的表现型非常复杂,即包括提高所有 gal 酶系的产量、对紫外线的高度敏感性,缺陷型细胞发生分裂(细胞要生长到几倍于正常细胞的长度时才发生分裂),以及产生粘性(细胞分泌多糖,致使菌落表面上附着一层有光泽的粘性物质),但是还不知道在 capR<sup>-</sup> 突变体中的分子缺陷是什么。

### 阿拉伯糖操纵子

阿拉伯糖是另一个可以为代谢提供碳源的糖(五碳糖)。在大肠杆菌中阿拉伯糖的降解需要三个基因: araB、araA 和 araD,它们形成一个基因簇,简称为 araBAD (见图 14-16)。这些基因分别编码三个酶: araB 基因编码核酮糖激酶(ribulokinase),araA 编码 L-阿拉伯糖异构酶(L-arabinose isomerase),以及 araD 编码 L-核酮糖-5-磷酸-4-差向异构酶(L-ribulose-5-phosphate-4-epimerase)。在半乳糖操纵子中,基因的排列顺序就是代谢途径中酶的作用顺序。这个操纵子却不同,ara 操纵子中的酶是以 araA、araB、araD 的顺序起作用的。另外还有两个负责将阿拉伯糖运入细胞里的基因,即 araE 和

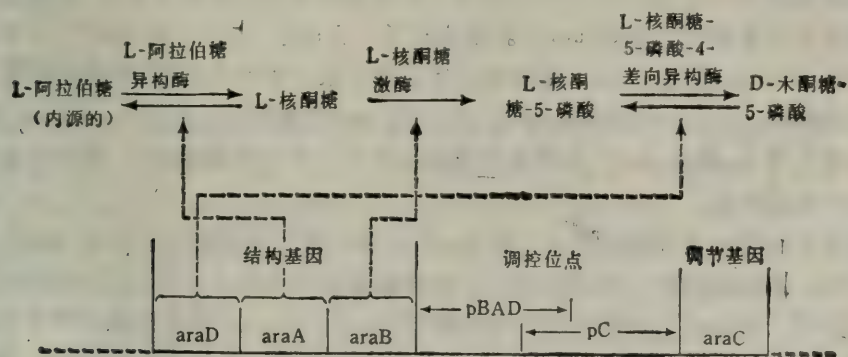


图 14-16 大肠杆菌的 ara 操纵子及其调控途径。pBAD 代表 ara BAD 基因的启动子, pC 代表 araC 基因的启动子;正文中将对这些启动子的重叠作出介绍。

\* “高水平”是一种通常的用法,用来表示高浓度或高速度(正确的含意通常要从上下文获得);反之,“低水平”用来表示低浓度或低速度。

araF, 它们位于远离这组基因的地方。araE 和 araF 分别地编码一个膜蛋白和一个在细胞壁与细胞膜之间的阿拉伯糖结合蛋白;但是目前我们对这两个基因还很不了解,所以不准准备在这里作进一步讨论。

与 araBAD 相邻的是一个复合的启动子区域和一个调节基因 araC, 这个调节基因的性质与我们已介绍过的 lac 及 gal 操纵子中的调节基因的性质全然不同。我们将看到 AraC 蛋白同时显示正负调节因子的功能。与之相反, 在 lac 和 gal 操纵子中, 阻遏蛋白只能作为负调节因子, 而 cAMP-CAP 蛋白只能是正调节因子。araBAD 和 araC 基因的转录是分别在两条链上以相反的方向进行的。在标准的遗传学图谱上, araBAD 基因簇从它的启动子 pBAD 开始向左进行转录, 而 araC 基因则是从 pC 向右转录。

ara 操纵子也是可诱导的, 阿拉伯糖本身就是诱导物。即在野生型操纵子中, 只有阿拉伯糖存在时才转录出 araBAD mRNA, 而在葡萄糖存在时则不转录。在腺苷酸环化酶的缺陷型和 crp<sup>-</sup> 突变株中, 细胞也都不能形成 araBAD mRNA。这些都说明了从 pBAD 起始的转录过程需要 cAMP-CAP。

### AraC 蛋白的正负调节作用

araC 基因产物有正调节活性, 以下结果可以说明这种活性:

1. 在 araC 基因内发生点突变或缺失突变, 产生出称为 araC<sup>-</sup> 的突变体。两种 araC<sup>-</sup> 都不能合成 araBAD mRNA。如果知道的仅此而已, 那么人们可能会认为或者可能是 araC<sup>-</sup> 细胞产生了突变型的阻遏物, 这种阻遏物, 不能与诱导物结合, 或者可能 araC 基因的位置就是启动子的位置。通过下面两项观察结果确认这两种可能性都不存在。

2. 基因型为 F'araC<sup>+</sup>/araC<sup>-</sup> 的局部双倍体的操纵子是完全可以诱导的。这意味着 araC<sup>-</sup> 等位基因是隐性的。假如一些 araC<sup>-</sup> 突变体的产物是不可诱导的阻遏物, 那么 araC<sup>-</sup> 对于 araC<sup>+</sup> 来说则应当是显性的, 然而这种看法与 2' 中观察到的实验事实不符。

3. 基因型为 F'araC<sup>+</sup>araA<sup>-</sup>/araC<sup>-</sup>araA<sup>+</sup> 的局部双倍体的操纵子对于合成 araA 产物来说是完全可诱导的。假如 araC 就是启动子, 又由于 araA<sup>+</sup> 与 araC<sup>-</sup> 突变处于同一个 DNA 分子上, 那么 araA<sup>+</sup> 基因就应当不被转录。事实表明 araC 基因产物互补于一个反式的构象 (trans configuration), 所以说 araC 基因不是启动子, 它与启动子并不在同一个位置上, 那么 araC 基因的产物则必然是一种可扩散的分子, 所以才能诱导合成出 araA 基因产物。

遗传学上还有一些其他的证据, 即 araC 基因上的一些琥珀突变说明 araC 产物是一种蛋白质, 这些遗传学结果由于 AraC 蛋白的纯化而得到进一步的证实;而且更进一步, 我们现在已经知道了该蛋白质的氨基酸序列。关于以上结果的实验细节我们没有必要去讨论, 关键问题是要懂得 araBAD mRNA 的合成需要一个有功能的 AraC 蛋白。

araC 基因产物对于从 pBAD 起始的转录也有微弱的阻遏作用, 这种看法是通过几个复杂的遗传学实验而得出的, 在这里不打算详细叙述。总的说来, 阿拉伯糖操纵子的一段区域可以缺失, 并且如果有 AraC 蛋白存在(因为它是 araBAD 转录所必需的), 在无阿拉伯糖时也可以产生一些 araBAD mRNA。这一现象使我们回忆起乳糖操纵子的操纵基因发生突变的情形, 虽然二者之间差异多于相似之处。例如, 比起有 lacO<sup>c</sup> 突变的



乳糖操纵子来说,阻遏作用是非常小的,  $\text{lacO}^c$  突变可以使操纵子基因完全表达,然而在  $\text{ara}$  操纵子中,当阻遏活性被排除时,产生的  $\text{araBAD}$  mRNA 的量仅是阿拉伯糖存在时(也就是说,当  $\text{ara}$  操纵子处于被诱导状态)的 1%。这种弱化阻遏作用的机理尚不清楚。我们即将看到  $\text{AraC}$  蛋白还有另外的阻遏活性,这次是对  $\text{pC}$  启动子的。

有两方面的体外实验进一步证实起始转录需要  $\text{cAMP-CAP}$  及  $\text{AraC}$  蛋白(当以正调控因子起作用时)的共同作用。在第一个实验里,当反应混合物中或者是(a)只含有  $\text{ara}$  DNA 和 RNA 聚合酶,或者是(b)含有 DNA、RNA 聚合酶以及蛋白质中的仅仅一种(只有  $\text{AraC}$  蛋白或只有  $\text{cAMP-CAP}$ )时,都不能合成  $\text{araBAD}$  mRNA,说明反应的进行需要两种蛋白质共同起作用。在第二个系列实验中,利用电镜检测  $\text{araDNA}$ (这种技术可以看到结合了 RNA 聚合酶的 DNA 分子),这些实验中的一个结果表明,除非同时存在着这两种蛋白质,否则 RNA 聚合酶是不能结合到 DNA 上的。后面我们将要这两种蛋白质是如何促进  $\text{araBAD}$  转录的进行一些解释。

### AraC 蛋白的两种形式

$\text{AraC}$  蛋白可作为  $\text{pBAD}$  活性正负调节子的双重作用,通过假设该种蛋白质具有两种异构体已经作出了一些解释。一种形式称为  $\text{P}_r$ ,是起阻遏作用的形式,可以与尚未鉴定的类操纵基因位点(operatorlike site)相结合。另一种  $\text{P}_i$  是起诱导作用的形式,它通过与  $\text{pBAD}$  启动子结合进行调节。推测  $\text{P}_r$  和  $\text{P}_i$  这两种结构处于一种相互平衡之中。在没有阿拉伯糖时,  $\text{P}_r$  形式占优势;一旦有阿拉伯糖存在,阿拉伯糖就能够与  $\text{AraC}$  蛋白结合使平衡趋向于  $\text{P}_i$  形式。这样,阿拉伯糖的诱导作用就可以解释为阿拉伯糖与  $\text{P}_r$  的结合使  $\text{P}_r$  离开它的结合位点,然后,产生大量的  $\text{P}_i$ ,  $\text{P}_i$  再与启动子结合。

关于  $\text{P}_i$  到  $\text{P}_r$  的转变还有许多未知的问题。例如,仍然不清楚阿拉伯糖是怎样诱导这种转变作用的。纯化的  $\text{AraC}$  蛋白的物理性质表明阿拉伯糖可以与它结合,并进行诱导,使它的构象发生变化。但其他一些实验指出  $\text{P}_i$  并不是一种简单的阿拉伯糖- $\text{AraC}$  蛋白复合物。

### AraC 蛋白合成的调节

$\text{AraC}$  蛋白的合成也是受到调节的,像  $\text{lac}$  操纵子中的  $\text{lacI}$  产生的阻遏物一样,它的合成过程也是受到自我调节的。即  $\text{AraC}$  蛋白本身作为  $\text{pC}$  的负调节子来控制其自身的合成过程,这种现象称为自我调节(autogenous regulation)。令人吃惊的是,  $\text{AraC}$  蛋白的合成也受到  $\text{cAMP-CAP}$  的调节。即葡萄糖存在时,没有  $\text{AraC}$  蛋白产生。因此,  $\text{araBAD}$  片段及  $\text{araC}$  基因的转录是协同调节的,这样做是比较符合经济原则的,因为没有其他几种蛋白质,仅仅单独地存在着一种蛋白质是没有意义的。

对  $\text{araC}$  基因产物调节机理的研究采用了一项新技术,从而避免了令人乏味的直接分析  $\text{AraC}$  蛋白的操作步骤。现在通过基因操作技术,将乳糖操纵子的  $\text{lacZ}$  基因从大肠杆菌中的正常位置上转移出来,并连接到  $\text{AraC}$  蛋白的启动子  $\text{pC}$  上(这被称为  $\text{lacZ-araC}$  启动子融合)。这种重排过的细菌不再具有  $\text{araC}$  基因(已被删除),但还是可以测到这个基因的转录活性;因为  $\text{lacZ}$  不再与自己原来的启动子偶联,而是处在  $\text{araC}$  的启动子  $\text{pC}$  的控制下,  $\beta$ -半乳糖苷酶仅仅由  $\text{pC}$  开始的 mRNA 上翻译出来。因此,所合

成出来的酶量(这是很容易检测的)就反映了 AraC 蛋白的合成。在测定大肠杆菌中经过基因融合的  $\beta$ -半乳糖苷酶的含量时,如果培养基中有葡萄糖,则酶产量降低,这说明araC 基因的转录需要 cAMP-CAP。如果把性质粒 F'ara 加到带有融合基因的菌株中,  $\beta$ -半乳糖苷酶的产量降低;但是,如果在这种携带 araC 的 F' 质粒上携带着 araC<sup>-</sup> 突变,则该酶的合成不受影响。因此, AraC 蛋白可以使  $\beta$ -半乳糖苷酶的合成减少。也由此推论, AraC 蛋白调节自身的合成, 而且这种调节必然是通过它与 araC 的启动子 pC 的结合而达到的。这些结果总结在表 14-5 中。

表 14-5 不同条件下 pC-lacZ 融合菌株中  $\beta$ -半乳糖苷酶的合成

基因型	葡萄糖	$\beta$ -半乳糖苷酶的合成	结 论
1.pC-lacZ	无	有	araC 的转录需要 cAMP-CAP
2.pC-lacZ	有	无	同 上
3.F'araC <sup>+</sup> /pC-lacZ	无	比“1”少	AraC 蛋白抑制自身合成
4.F'araC <sup>-</sup> /pC-lacZ	有	与“1”同	同 上

注意, lacZ 基因从 araC 的启动子 pC 开始转录。融合作用中不存在 araC 基因。

### ara 操纵子中调节结合位点的定位

关于阿拉伯糖操纵子的调节过程已经知道得很多,其中关键性的信息就是一些蛋白质结合位点的位置(图 14-17)。这些位点的确定需要借助于序列分析。首先,测定整个 ara 调控区域的碱基序列。这个区域从 araB 基因内部起一直延伸并穿入 araC 基因。然后,再利用核酸酶-保护技术(在本书中我们已多次提到过这个技术)分别确定存在着阿拉伯糖或缺乏阿拉伯糖时被 cAMP-CAP 蛋白和 AraC 蛋白保护的以及被 RNA 聚合酶保护的位置。转录的起始位点则通过测定 mRNA 的序列而得知。其他一些实验指出 P<sub>r</sub> 和 P<sub>i</sub> 在结合到 pC 上时与 RNA 聚合酶竞争;而且,当 RNA 聚合酶结合到 pBAD 上时同时还需要有 cAMP-CAP 及 P<sub>i</sub> 参与。图 14-17 表示了调节区的有关部分,共有五个重要的区域,下面对这些区域以及它们的作用作一些说明:

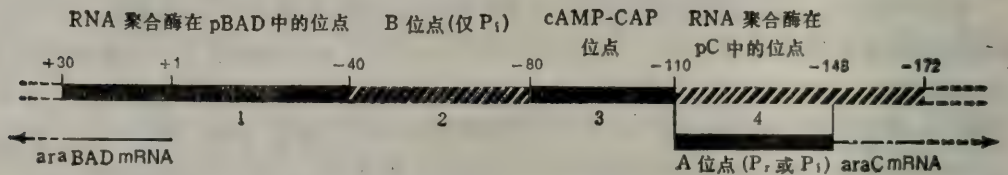


图 14-17 ara 操纵子的调节区。示出 RNA 聚合酶的结合位点、cAMP-CAP 以及 AraC 蛋白的 P<sub>r</sub> 形式和 P<sub>i</sub> 形式。数字代表核苷酸数; araBAD mRNA 的起始点用数字“1”表示。A 位点全部包含在 pC 上的 RNA 聚合酶的作用位点之内。

1. RNA 聚合酶在 pBAD 中的结合位点 1,它与对 araBAD 序列进行转录有关。

2. B 位点 2,它会对 P<sub>i</sub> 所参与的 araBAD 转录过程的正调节作出反应。这个位点的重要性在于除非它与 P<sub>i</sub> 结合(即被 P<sub>i</sub> 占据),否则 RNA 聚合酶就不能与 pBAD 结合。



3. cAMP-CAP 结合位点 3, 这是一个 pBAD 和 pC 共用的位点。此位点调节它左面的 araBAD 基因及右面的 araC 基因的表达。当这个位点没有结合到 cAMP-CAP 时, RNA 聚合酶对 pBAD 和对 pC 的结合也就因此而减弱。

4. RNA 聚合酶在 pC 中的结合位点 4, 该位点与 araC 序列的转录作用有关。

5. A 位点, 这个位点处于位点 4 之内, 被认为是 araC 基因的操纵基因。P<sub>r</sub> 和 P<sub>i</sub> 都可以与 A 位点结合, 在这样的情况下抑制 RNA 聚合酶与 pC 的结合。因此, 如果 AraC 蛋白过量时, 它的合成就被降低; 但是, pC 的转录并不会因为 P<sub>r</sub> 及 P<sub>i</sub> 而全部关闭。这样, 只要存在着阿拉伯糖, 总是有足够的 P<sub>i</sub> 去占领 B 位点, 使得从 pBAD 起始的转录作用能持续下去。

注意 pBAD 启动子包括位点 1, 2 和 3。然而 pC 启动子则由位点 3 和 4 组成。

### 细胞营养状态对 ara 操纵子的影响

现在我们来观察在不同营养条件下这些调节区域的状态 (见图 14-18)。在图 (a) 中, 存在着葡萄糖, 所以在 cAMP-CAP 的结合位点上没有 cAMP-CAP 结合。因为缺乏阿拉伯糖, 所以 AraC 蛋白处于 P<sub>r</sub> 形式并与 A 位点结合。这时, A 位点被 P<sub>r</sub> 占据, 因此 RNA 聚合酶很少再与 pC 结合, 于是几乎不再形成 AraC 蛋白。整个系统都处于静止状态, 因为在有葡萄糖时它没有必要处于激活形式。它虽然不是完全地关闭 (永远不会完全关闭), 但是尽可能地关闭了。在 (b) 中没有葡萄糖, 也没有阿拉伯糖。因为

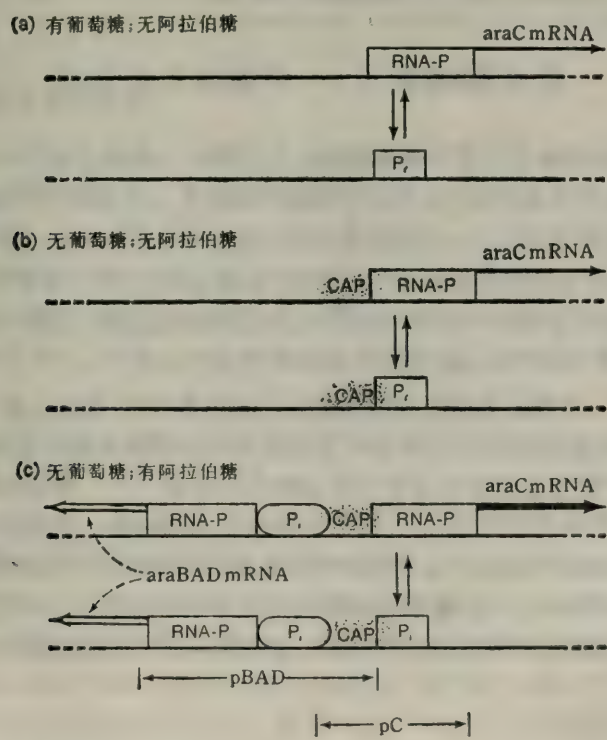


图 14-18 ara 操纵子的调节区处于各种营养条件下的状态。RNA-P 代表 RNA 聚合酶, CAP 代表 cAMP-CAP。

cAMP 的量有所提高, cAMP-CAP 位点被占据。 $P_r$  仍使 AraC 蛋白的浓度保持在低水平,但由于它不能以绝对优势与 RNA 聚合酶竞争(其竞争效率不如  $P_i$ ),所以合成出稍多一些 AraC 蛋白。除非有其他碳源存在,否则这种处于葡萄糖饥饿状态下的细胞是不能生长的。随着 AraC 蛋白浓度提高,细胞处于一种准备状态,一旦供给阿拉伯糖, $P_r$  将马上转变为  $P_i$ ,并且把 B 位点完全占据了。(c) 具有阿拉伯糖而没有葡萄糖的情况。这是完全被诱导的状态,B 位点和 cAMP-CAP 位点都被占据了,这时 RNA 聚合酶就能够与 pBAD 结合,从而产生 araBAD mRNA。A 位点也被  $P_i$  结合, $P_i$  能够与 RNA 聚合酶有效地竞争 pC 启动子。这样就避免了过多地合成 AraC 蛋白所造成的浪费。但当这种蛋白质( $P_i$ )的浓度降低到不足以维持从 pBAD 的转录时,A 位点就空出来, RNA 聚合酶重新进入 pC,从而产生 araC mRNA。

如果把葡萄糖加入到已完全被诱导的体系中,就不再需要这些 ara 酶了。葡萄糖降低 cAMP 的浓度,因此 cAMP-CAP 位点将不被占据, RNA 聚合酶与 pBAD 的结合也随之减少,当然也就减少 ara 酶的合成。这种经济的作法是符合细胞的利益的。与此相似,当阿拉伯糖耗尽时(也没有葡萄糖存在) $P_i$  将转变成  $P_r$ ,B 位点空出来,这时 araBAD mRNA 的合成也成为一种浪费,这种合成就被阻止了。

注意阿拉伯糖操纵子与乳糖操纵子及半乳糖操纵子的不同之处是这个体系有两个正调节子( $P_i$  和 cAMP-CAP),它们串联着,先后地起作用激活同一个启动子(pBAD)。在这个区域(B 位点和 cAMP-CAP 位点)中 cAMP-CAP 使转录向两个方向进行(分别从 pBAD 和从 pC 出发,进行左向转录和右向转录)。

## 色氨酸操纵子,一个生物合成体系

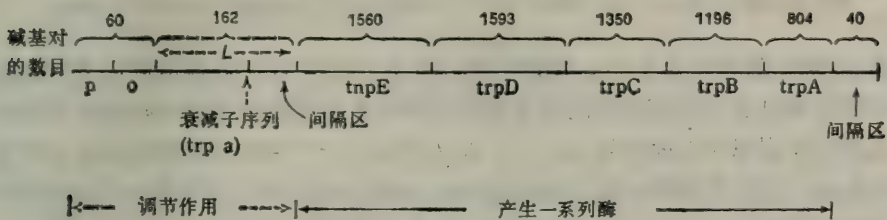
色氨酸(trp)操纵子负责合成色氨酸。这个操纵子的调节基于一个简单的原理,即如果向培养基中加入色氨酸,便没有必要再激活操纵子。由此我们可以想到,在有适量的色氨酸存在时调控体系关闭 trp 基因的转录;而在没有色氨酸时又打开 trp 基因的转录。换言之,色氨酸,或与此有关的某种物质,将在阻遏过程中(而不是诱导过程中)起作用。此外,由于 trp 体系形成一个生物合成的而不是降解的途径,它将不受葡萄糖的抑制作用;事实上 cAMP-CAP 对 trp 操纵子的活性不起什么作用。

在一般情况下,细胞总是能够得到少量的外源性色氨酸,但是,如果细胞自己的色氨酸合成体系完全关闭,那么这些外源供给就往往不足以维持细胞的正常生长,在这种情况下细胞可以通过两种手段避免自身处于色氨酸饥饿状态:(1)直到外源性色氨酸的浓度超过一定阈值时,阻遏作用不发生,才能产生合成色氨酸的酶。然而这种手段不是特别有效,因为平时经生物合成途径产生的酶往往已经超过需要量了。(2)建立一个调节体系,使色氨酸的浓度可以决定 trp 基因在脱阻遏状态下的转录量。这个机理比前一个有效,而且也是在 trp 操纵子中实际发挥作用的一种调控手段(其他氨基酸生物合成的操纵子与此相同)。



## trp 操纵子的组成

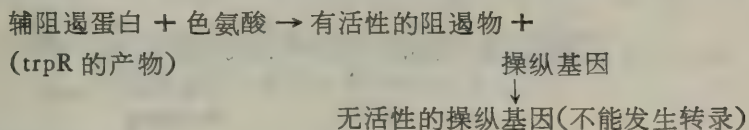
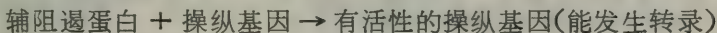
色氨酸生物合成过程有五个步骤，每一步都需要一个特殊的酶。大肠杆菌中编码这些酶的基因的排列顺序与这些酶在合成途径中的作用顺序相同。它们是从一条多顺反子的 mRNA 上翻译下来的。这些基因分别称为 *trpE*, *trpD*, *trpC*, *trpB* 和 *trpA*。最先翻译的是 *trpE*, 与 *trpE* 相邻是启动子、调控基因和两个分别被称为前导序列 (leader) 及衰减子 (attenuator) 的区域。这两个区域分别用 *trpL* 和 *trpA* (注意不是 *trpA*) 表示 (见图 14-19)。阻遏物基因 *trpR* 远离这一基因簇, 就像半乳糖操纵子中的阻遏物基因一样。色氨酸操纵子中还有两个因子参与调控作用, 即 tRNA<sup>Trp</sup> 和色氨酸-tRNA 合成酶, 编码这两个因子的基因簇也相距甚远。



**图 14-19** 大肠杆菌的 *trp* 操纵子。为清楚起见,将调节区放大(相对于编码区而言)。实际上每一个区域的合适的大小已用碱基数标出。L 是前导序列,调节区域用虚线表示。

### trp 的阻遏蛋白-操纵基因体系

与乳糖操纵子一样,如果在  $\text{trp}$  操纵子的  $\text{trpR}$  基因中和在操纵基因中发生突变,将起始组成型  $\text{trp mRNA}$  的合成。除非在色氨酸存在下,否则  $\text{trp}$  基因的蛋白产物是不会与操纵基因结合的。 $\text{trpR}$  基因的蛋白产物常常被称为  $\text{trp}$  的辅阻遏蛋白。辅阻遏蛋白(aporepressor)与色氨酸在一起则形成一个有活性的阻遏物,这个有活性的阻遏物可以再与操纵基因结合,使操纵基因失去活性,进一步抑制转录。这一反应可图示如下:



如果停止外源性色氨酸的供应(或大幅度降低供应),以上反应式的平衡向左移动,操纵基因的位置空出来,又开始转录。这就是基本的开启-关闭调节机理。

启动子与操纵基因的区域有很大程度的重叠,有活性的阻遏物的结合与 RNA 聚合酶的结合发生竞争。在体外时,如果先加入阻遏物,则 RNA 聚合酶不能结合,反过来也一样。至于阻遏物在这里是否像 Lac 阻遏物那样作为一个抗融解蛋白 (antimelting protein) 起作用,则还不是十分清楚。

图 14-20 表示启动子-操纵基因及与其相邻的区域的碱基序列, 这个相邻区包括

trp mRNA 合成的起始位点。操纵基因的20个碱基对中的18个碱基对形成了倒转重复的对称结构。

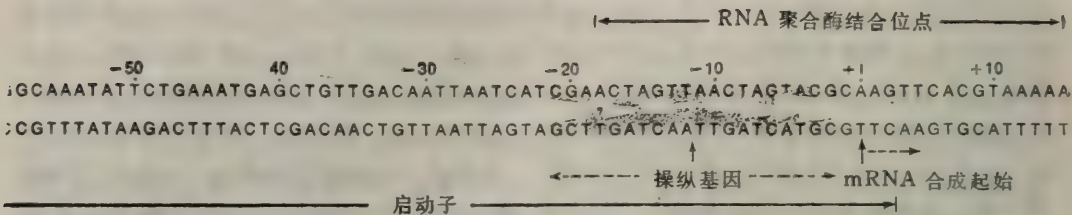


图 14-20 trp 操纵子的启动子-操纵基因的碱基序列。阴影区表示操纵基因,对称中心为操纵基因上方的箭头所指之处。trp 操纵子中第一个基因的 AUG 起始密码子在+162 位置处。

阻遏物-操纵基因形成了 trp 操纵子的一个有效的“开启-关闭”开关。然而，在氨基酸生物合成的操纵子中还有另外的更为精细的调控机理使酶浓度能够随着氨基酸浓度而变化。这种调控来自两方面的效应：（1）在转录到第一个结构基因前时就发生转录作用的早熟性终止；（2）氨基酸的浓度控制这种终止转录的频率。在下面一节里我们将详细地描述这个过程。

### 衰减子和前导多肽

从 trp mRNA 的 5' 端开始,在 trpE 基因的起始密码子的前面有 162 个碱基(见图 14-19 的左部)。这一段 mRNA 称之为前导序列（对这个区域的一般性描述见第十一章）。如果在先导序列中删去一段 DNA（第123—150 碱基对),那么不论在脱阻遏的还是在组成型的突变细胞中,这些 trp 酶的产量都将增加约 6 倍。所以,123 到 150 这段序列必定具有调控活性。而且发现,在 mRNA 合成起始以后,除了在完全没有色氨酸的情况下以外,大多数 mRNA 分子都将在这—区域终止,从而产生一个具有仅仅 140 个核苷酸的 RNA 分子,即在编码 trp 酶的基因还没有被转录之前就终止了转录；这就解释了为什么一旦这个区域发生缺失,就会增加基因表达。在这个转录作用终止并受到调节

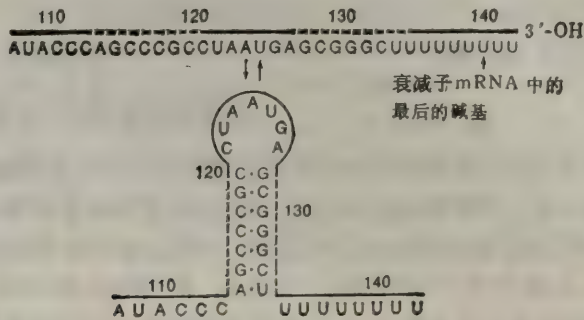


图 14-21 trp 衰减子 mRNA 的终止区。图中所示的碱基序列一直延伸到终止位点,终止位点处于 mRNA 中的 140 处,是一连串 U 组成的序列。图中用虚线标出的碱基形成倒转重复序列,这种倒转重复序列可以造成图中下部所示的柄与泡结构。



的区域被称为衰减子 (attenuator)。仔细地分析终止发生的这个区域的碱基序列 (见图 14-21), 发现它具有终止位点的一般结构, 即如第十一章中所介绍的一个柄与泡的结构 (stem-and-loop) 以及随之而来的一连串为五个碱基的 A.T 序列。

来自各方面的实验表明衰减作用需要有带负载的  $tRNA^{Trp}$  参与, 这意味着前导序列的某些部分被翻译了。分析前导序列, 发现它包括一个 AUG 密码子, 随后还有一个同时协调的 UGA 终止密码子; 如果翻译起始于 AUG, 将产生一个含有 14 个氨基酸的多肽 (见图 14-22)。这个假设的多肽 (还未实际观察到) 被称为前导多肽 (leader peptide)。有一些证据表明这个肽可能实际上是存在的。例如, 刚好在 mRNA 上的 AUG 位点的 5' 一侧中有一个核糖体的结合位点。此外, 前导肽的编码区与 *trpE* 基因末端的融合能够产生一种融合蛋白质, 该蛋白质具有与前导肽同样的 N 末端序列。

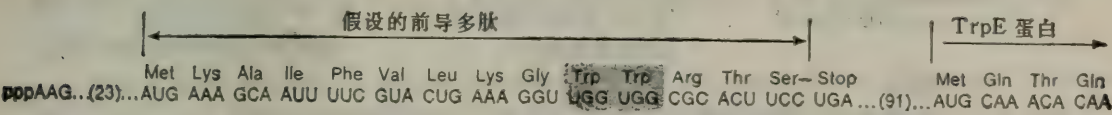


图 14-22 *trp* 操纵子的前导 mRNA 碱基序列和假设的前导多肽序列。图中还指出了两个编码色氨酸 *trp* 的密码子 (阴影部分) 和 *TrpE* 蛋白质的起始部分。为了清晰起见, 删去的碱基数用数字 (23 及 91) 表示。

前导多肽有一个非常有意思的性质, 即在第 10 和 11 位上有两个紧靠着的色氨酸密码子。与组氨酸操纵子比较, 我们可以看到这一性质的重要意义。在组氨酸操纵子中同样也有一个衰减子体系 (即有一个尚未完全成熟就被终止合成的 mRNA), 并且也包含着一个相似的编码前导多肽的序列。这个序列具有 7 个紧靠着的组氨酸密码子 (见图 14-23), 我们接着就要讨论这些重复密码子的意义。

- (a) Met Lys His Ile Pro ~~His His~~ Phe Ala ~~Phe Phe Phe~~ Thr ~~Phe~~ Pro Stop  
5' AUG AAA CAC AUA CCG UUU UUC UUC GCA UUC UUU UUU ACC UCC CCC UGA 3'
- (b) Met Thr Arg Val Gln Phe Lys ~~His His His His His His His~~ Pro Asp  
5' AUG ACA CGC GUU CAA UUU AAA CAC CAC CAU CAU CAC CAU CAU CCU GAC 3'

图 14-23 (a) 苯丙氨酸操纵子和 (b) 组氨酸操纵子的前导多肽的氨基酸序列和 mRNA 相应部分的碱基序列。重复的氨基酸用阴影表示。

mRNA 前导序列中的碱基配对

*trp* 前导序列的碱基顺序已测定出来, 其中引人注目的是其中的四个片段用 1、2、3 和 4 代表, 它们能够以两种方式进行碱基配对 (见图 14-24), 即可以以 1—2 和 3—4 或只是 2—3 的方式在它们之间形成互补配对。利用 RNase T1 (此酶不能水解配对的 RNA) 的作用表明纯化的 *trp* 前导 mRNA 中只有 1—2 和 3—4 的配对方式。因为由这样定位的 3—4 配对区应当处于终止密码子的识别区域之内, 所以它的确是转录终止所必需的, 即在这一区域内发生的几种突变, 或是在体内实验中阻止了转录的终止, 或是在体外实验

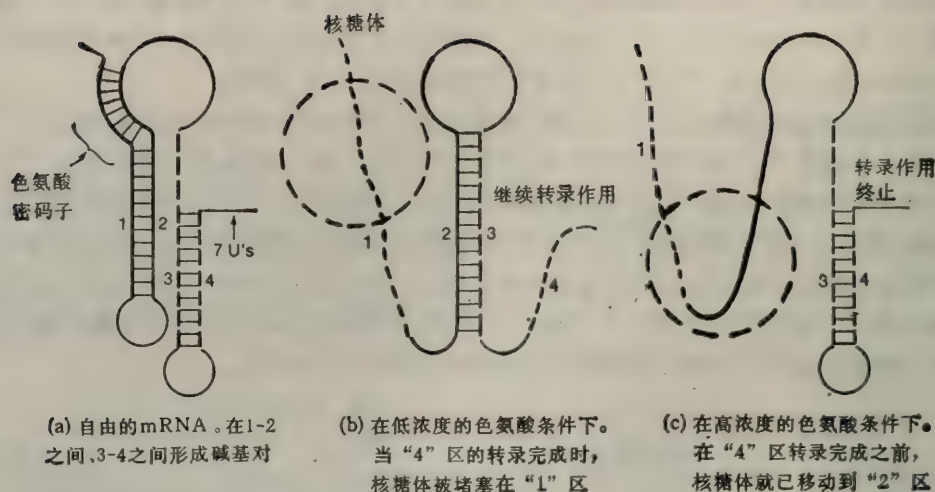


图 14-24 已被人们接受的大肠杆菌 *trp* 操纵子中的衰减作用机理的模型。

中降低了这个区域对 RNase T1 酶的抗性。值得注意的是，由于双螺旋的 1—2 片段和 3—4 片段中的 2 与 3 序列相互配对，于是有 2—3 区，而不可能同时存在 1—2 区和 3—4 区。这也就意味着如果条件适宜于形成 2—3 区，那么 1—2 区和 3—4 区便不复存在，这是很重要的一点。从以上这些现象中可以总结出关于 *trp* 前导序列过早终止转录的机理，这一理论虽然尚未得到严格证明，但已被广泛地接受，我们将在下一节讨论。

### 衰减的机理

这一理论(见图 14-24)提出前导多肽“基因”的翻译促进了 mRNA 的终止。因为这个基因里有两个色氨酸密码子，这段序列的翻译能力将对携带了色氨酸的  $tRNA^{Trp}$  的浓度敏感，即如果色氨酸的量是有限的，那么就会使带有色氨酸的  $tRNA^{Trp}$  不足，从而穿过色氨酸密码子的翻译作用的速率就应当非常低。这一理论还推测像细菌中的一般情况那样，转录和翻译过程是偶联在一起的，而且，在 mRNA 与核糖体接触的区域中根本没有碱基配对。图 14-24 中指出 *trp* 前导肽的末端位于第一区。通常，进行翻译的核糖体越过那些正被翻译的密码子与 mRNA 中的大约 10 个碱基相互接触，这样当最后一个密码子被翻译完毕时，第 1 区和第 2 区还不能配对。在转录-翻译的偶联体系中，前导的核糖体不会远远地落在 RNA 聚合酶之后。因此，当第 4 区正在完成其转录合成时，核糖体就与第 2 区接触；然后，第 3 及第 4 区可以自由地形成 3—4 区双螺旋区，而不会有第 2 区为了形成双螺旋而产生的竞争。3—4 区的柄与泡结构的存在，使得 RNA 聚合酶一旦达到 7 个 U 的终止序列时，转录作用立即终止。如果不再加入色氨酸，有负载的  $tRNA^{Trp}$  的浓度降低到一定的限度。正在翻译着的核糖体将偶尔地在色氨酸密码子处滞留(这些密码子位于第 2 区开始前的 16 个碱基处)。因此，在第 4 区被合成之前，第 2 区可以自由地形成 2—3 区的双螺旋。在没有 3—4 区的柄与泡结构的情况下，转录不能终止，从而可以把包括所有的 *trp* 基因在内的序列转录成一个完整的 mRNA 分子。所以如果色氨酸过量，就使转录作用终止，几乎不产生酶；如果没有色氨酸，则转录不终止，所有的 *trp* 酶都能合成。



## 阻遏作用与衰减作用的对抗

trp 的阻遏物-操纵基因体系不能像“开启-关闭”开关那样地起作用，但是能够使细菌产生中等浓度的色氨酸。人们已经知道，在不加色氨酸的条件下，在细胞生长的全部时间里 trp mRNA 的合成都受到部分阻遏，因为在一个阻遏物发生突变的(无活性)细胞中 trp 酶系的浓度要比一个野生型细胞中该酶系的浓度高出 10 倍。人们观察到的上述现象表明，如果内源的色氨酸浓度由于任何原因引起起伏不定，那么野生型细胞中有活性的阻遏物与无活性的阻遏物之间的平衡就会倾向到能使色氨酸有较为方便来源的那一侧，使色氨酸维持在一定的水平上。因此，人们现在还没有完全弄清楚为什么还需要衰减子体系。

有可能，有活性的阻遏物向无活性的阻遏物的转变速度极低，这种极低的转变速度就需要存在一种衰减体系。这种衰减体系能更迅速地作出反应，使色氨酸能从较高的浓度迅速地下降为中等浓度。

也有可能，在有外源性色氨酸存在时需要有关衰减体系。如在外源性色氨酸浓度实在太低，生物本身又没有另外的内源性的色氨酸合成体系，细菌甚至无法支持其自身的生长时，需要有衰减体系加以调节。例如，外源性色氨酸的浓度很低的信号对于增强阻遏作用已是足够的了，但是这个信号的强度还不足以像合成一些必需蛋白质时那样能触发和加强内源性合成色氨酸的过程。于是，在这种环境下，衰减子就应当对总的较低的内源性色氨酸的浓度作出反应——通过不终止(即抗终止) mRNA 的合成来增加 trp 酶系的合成从而提高内源性色氨酸的浓度。在这种情况下，当内部色氨酸浓度由于环境影响而变动很大时，衰减子可以被看作是一个对这种变动起调控作用的调节器。

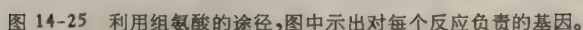
人们可能会问为什么要有阻遏体系；没有阻遏体系，似乎只有衰减子也可以调节色氨酸酶系的合成。阻遏物的作用可能是当有大量外源性色氨酸存在时，阻止非必需的先导 mRNA 的合成，它使这个合成系统更加经济。事实上，自然界中存在着不同类型的合成体系，例如组氨酸操纵子，它拥有在功能上与色氨酸操纵子中的衰减子完全相同的衰减子，这个衰减子能够形成一个有三个碱基对的区域，但组氨酸操纵子没有阻遏物，这说明完全只是受衰减子调节的。

目前，我们不知道为什么色氨酸操纵子具有双重调节体系，而组氨酸操纵子却不具有双重调节体系，这些都需要作进一步的探讨。

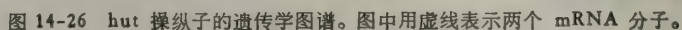
许多负责合成氨基酸的操纵子都像色氨酸操纵子那样，是受到衰减子调节的。这些衰减子都有碱基配对机理，就像色氨酸操纵子里的竞争配对那样。虽然现在人们对于大肠杆菌、鼠伤寒沙门氏杆菌和沙雷菌属的粘质赛氏杆菌 (*Serratia marcescens*) 中组氨酸、苏氨酸、亮氨酸、异亮氨酸-缬氨酸以及苯丙氨酸等的操纵子的了解，与对色氨酸操纵子相比，所了解的细节较少，但也已经见到不少对它们的介绍。

## 利用组氨酸的操纵子

许多氨基酸除了能被用来合成蛋白质以外，还可以在碳源或氮源不足的时候，作为碳原子或氮原子的来源维持细胞的生长。由于这种双重作用，氨基酸的降解必须受到严



hut 操纵子共编码四种酶和一个阻遏物。四种酶分别由 hutG、hutH、hutI 及 hutU 基因编码,阻遏物则由 hutC 基因编码。在产气克雷伯氏菌 (*Klebsiella aerogenes*) 中,以上这些基因组成为两个转录单位,即含 hutI、hutG、hutC 的单位和含 hutU、hutH 的单位(图 14-26)\*\*。这两个转录单位各自都有一个启动子和一个操纵基因,并如图所示,其转录过程都是从左向右进行的。推测两个操纵基因是相同或近似于相同的, hutC 阻遏物分子与每一个操纵基因结合。虽然,实际上的诱导物是组氨酸被 hutH 的基因产物作用而形成的产物尿刊酸盐(urocanate),但是,诱导过程的完成要靠加入组氨酸。尿刊酸盐与阻遏物结合,将阻遏物转变成无活性形式,使合成 mRNA 的每个启动子能处于自由状态。然而,由于像 lac 操纵子那样,也需要有一个正调节子,所以这时仍然没有转录出 hut mRNA。hut 操纵子的一个有趣之处,就是它有两种不同的正调节子(像 ara 操纵子一样),但是对于 RNA 聚合酶结合到启动子上的过程,只需要其中的一个正调节子。



最近,这些区域已被充公并分离得到,这对于弄清一些很多年来未清楚的问题来说,是一个巨大的进步。



方式与谷氨酰胺合成酶的合成相关联(谷氨酰胺合成酶负责将氨用于合成含氮化合物)。最近,已分离到几种氮调节体系的组分,并且已经建立了 hut 操纵子的体外转录体系。因此,不久将能更加清楚地了解 hut 体系的功能。

表 14-6 总结了 hut 操纵子的主要活性。

请注意,许多教科书叙述了谷氨酰胺合成酶在感受氮量体系中的直接作用以及加入 AMP 后就失去活性。最近已经证明这些看法是错误的。

表 14-6 在不同培养基条件下 hut 操纵子活性的总结

蔗 糖	组氨酸	氮 源	碳 源	cAMP*	操纵子的活性†(“开”或“关”)
—	无	—	—	—	关
有	有	充足	充足(来自蔗糖)	低	关
无	有	充足	只来自组氨酸	高	开
有	有	只来自于组氨酸	充足(来自于蔗糖)	低	开
无	有	只来自于组氨酸	充足(不来自于蔗糖)	高	开

注意:横线表示该项无关紧要。

\* 碳源与 cAMP 浓度的联系是:高 [C]→低[cAMP]和低 [C]→高 [cAMP]。

† 只要 cAMP 是处于高浓度时,操纵子就是有活性的。

recA 操 纵 子

在第九章介绍了 SOS 修复系统,这是一个可诱导的体系,当 DNA 遭到严重破坏时,细菌中这个体系就会被启动。这个可诱导体系的一个主要产物就是 recA 基因的产物。这个蛋白质在修复与遗传重组(对于细菌,这些过程是十分重要的)中起作用,它有好几种活性:它能与单链和双链 DNA 结合,它具有 ATP 酶的活性,同时它还是一种蛋白酶。对大部分蛋白质起作用时这种蛋白水解酶的活性是很弱的,但它对噬菌体中特定的阻遏物的水解活性却很强,它将在一个独特的位点把每一个阻遏物切开。不久,我们将转向讨论这个重要之点。细菌被紫外线照射其 DNA 受到损伤之后,突然大量地合成 recA 的 mRNA。但是,如果 recA 突变成 recA<sup>-</sup>,这种合成就不会发生,这与至此所介绍过的各种操纵子中所观察到的情况大不相同。例如,在一个具有 galK<sup>-</sup> 的细菌突变株中加入半乳糖,虽然它合成出的 galK 产物是有缺陷的,是没有酶活性的,但是半乳糖仍然能够开动这个 gal 操纵子。对于 recA 操纵子,则必须有有功能的 RecA 蛋白才能启动这个操纵子。对于这种自控活性 (autoregulating activity) 解释为: RecA 蛋白像 araC 产物一样,是一个正调节子,另一种解释为 recA 的基因产物具有诱导活性,我们将会看到后一种解释是正确的。

recA 操纵子中的另一个组分是 lexA 基因,lexA 可能编码操纵子的一个简单的阻遏物。已分离到一些 LexA<sup>-</sup> 的突变体,其表现型为 RecA<sup>-</sup>, 这些突变体显然是不可诱导的,因为它们受到紫外线照射后不能作出反应,即不能相应地增加 RecA 蛋白的合成量。

许多遗传学的证据以及体外生化试验的证据都指出,在未受损伤的细胞中 RecA 蛋

白只有结合 DNA 的活性而没有蛋白水解酶的活性;反之,纯化了的 RecA 蛋白却具有蛋白水解酶的活性。所以人们认为,这种体外观察到的现象与体内观察到的结果的区别是因为在蛋白质分离过程中或是有活化作用或是丢失了一个抑制子 (inhibitor)。目前认为,在 DNA 损伤后会发生以下两种情况中的一种: (1) RecA 蛋白结合到在修复早期中所产生的 DNA 单链区上,从而转变成为有活性的蛋白酶,推测这可能是由于构象改变造成的; (2) 某些是修复早期的一些副产品,即某些效应物分子,对 RecA 蛋白进行活化。目前关于这两点孰是孰非还在广泛地进行讨论。

已分离到 LexA 蛋白。在体外,纯化了的 RecA 蛋白将 LexA 蛋白切成两个片段。已经分离得到一种突变株 LexA3 的基因产物。它不具有 LexA<sup>+</sup> 细菌的特点,也就是说它不能增加 RecA 蛋白的合成量,体外实验指出, LexA3 蛋白对于纯化了的 RecA 蛋白对它的切割作用有抗性。所以,纯化了的 RecA 蛋白不能切开 LexA3 蛋白。

上述的这些实验事实和理论可以对 recA 操纵子的调控作出解释。正常情况下,正在生长的细胞含有组成型水平量的 RecA 蛋白,这种组成型 RecA 蛋白没有水解蛋白酶的活性。在 DNA 受到损伤之后,通过一种未知的机理这种少量的 RecA 蛋白分子被激活,转变成有蛋白水解酶活性的形式。然后, RecA 蛋白水解酶切断所有的 LexA 阻遏物分子。因此, recA 操纵子转变为脱阻遏状态, RecA 蛋白的合成就增加了。只要活化信号还存在着,操纵子就一直处于活化状态,但是当修复完成后,活化信号就立刻消失, RecA 蛋白又回到非蛋白水解酶的形式。一直连续不断地合成着的 LexA 蛋白,仅仅在出现这种形式时,才能积累起来,并重新建立阻遏作用。 RecA 蛋白的脱阻遏的合成停止,细菌又开始生长并进行分裂, RecA 蛋白又逐渐地稀释到组成型的水平。

## 核糖体合成的调节

在前一节中已介绍了一些操纵子,它们的调控作用已研究得很清楚,这里我们将要讨论一个很复杂的调控体系,这个调控体系已研究了 30 年,虽然已掌握了许多信息,但是它的机理对于我们仍然还是一个谜。

所有各个种类的细菌的生长速度都会因培养基不同而不同的。在含有足够碳源 (如蔗糖) 的培养基上,大肠杆菌在 37℃ 下每 45 分钟分裂一代;在碳源供应较差如给以脯氨酸时,则要经过 500 分钟才分裂一代。在极为丰富的含有蔗糖、氨基酸、核酸碱基、维生素和脂肪酸的完全培养基上,细菌不必合成这些物质,因此能快速地生长;典型的情况下,在这种丰富培养基上,细菌大约每 25 分钟就分裂一次。细菌能有这些不同的生长速度是通过蛋白质合成的能力加以调节而达到的。这种调节作用可以通过改变以下几个方面而完成: (1) 因加入氨基酸肽链延长的速度不同; (2) 起始一条新的肽链分子的速度不同。事实上,肽链的合成速度是固定的,蛋白质合成的总速度是由每个细胞中的核糖体数目决定的。在快速生长的细胞中,细菌的分裂晚于 DNA 的合成,因此快速生长的细胞中的 DNA 多于生长较慢的细胞中的 DNA。例如,表 14-7 所指出的在三种不同生长速度的细胞中每个细胞中的 DNA 分子数不同。还应当注意到表中所示的细胞中核糖体数目与 DNA 分子数目虽然并不成比例,但是也随着生长的加快而增多。进一步还发现 tRNA、翻译起始因子、翻译延长因子以及氨基酰-tRNA 合成酶等的数目也都与表 14-7



表 14-7 不同生长速度的大肠杆菌的一些特点

分裂时间(分钟)	每个细胞中的 DNA 分子数	与每个 DNA 分子相关的核糖体数
25	4.5	15 500
50	2.4	6 800
100	1.7	4 200
200	1.4	1 450

中每种 DNA 分子数目相关的核糖体的数目有关。因此,人们认为蛋白质生物合成过程中的核糖体蛋白质、rRNA 以及蛋白质合成装置中的其他一些组分的合成也都是受调节的。

以上这些物质,至少其中一部分的合成过程是通过被组构到操纵子中而受到调节的。基因定位图谱实验指出,大肠杆菌中编码 53 个核糖体蛋白质的基因都位于染色体的一个小区段中,但是这些基因不是一个一个地转录,而是被转录成大约 20 个多顺反子 mRNA。rRNA 的基因分布在其中的三个几乎相同,但不连在一起的区域中,这些区域靠近核糖体蛋白的基因。在这些 rRNA 的基因簇中,如第十一章所述,也包含有某些 tRNA 基因;剩下的其他 tRNA 基因则位于分散在整个染色体上的几个基因簇中。目前还不清楚这些 DNA 片段是否为协同调控的,但看来是不可能的,因为人们还可以作出以下更为合理的设想:一个操纵子的产物调节着另一个操纵子,因为在这种调控方式中,各个操纵子可以产生出有恒定比例的产物。实际上,编码核糖体蛋白质的 6 个 mRNA 分子的翻译过程都是由几种特定的核糖体蛋白质(每一个多顺反子 mRNA 提供一种核糖体蛋白质)的浓度自动调节的。因此,受到控制的核糖体的装配速度,是部分地由 rRNA 的合成速度来调节的;而核糖体的装配又反过来间接地调节了这些蛋白质的生物合成过程。

如果将细菌培养物从快速生长的培养基中转移到一种让细菌只能慢速生长的培养基中(这种降速生长被称作“下移”,downshift)时,每一个细胞中的核糖体数目就会减少。由在快速培养基中的较高水平下降到在慢速培养基中才具有的特征性的较低水平。这是合理的,否则的话,每个细胞中就会有大于它所需要量的核糖体。只要让细菌继续地进行 DNA 的合成,而不允许合成 rRNA,细菌就可以完成上述降速生长过程。每个缓慢生长着的细胞必定是感受到某些信号,而且这些信号必然抑制 rRNA 基因的转录。在下移时,由于提供的碳源减少,所以在慢速培养基中,氨基酸的合成要比在快速培养基中慢。有几个证据暗示出这种氨基酸浓度的减少就是刚才所提到的信号之一。将需要某氨基酸的突变菌培养物从含有这种氨基酸的培养基上转移到具有相同碳源但无该氨基酸的培养基中,这样就可以模拟出降速生长的效果。事实上,在转移之后,rRNA 的合成很快就停止了,这种随着氨基酸的耗尽而抑制 rRNA 的合成的联系被称作是严紧反应(stringent response)。人们通过研究一些突变株,逐渐地明白了负责严紧反应的一些因子,这些突变株在氨基酸饥饿时,丝毫不减弱合成 rRNA 的过程。这些突变株的表现型被称为是松弛型(relaxed)。在几个基因的突变图谱中这些突变的位置被记作 rel。在 rel 基因中完全弄清楚的唯一基因就 rel A。

在氨基酸饥饿时,会产生两种不同寻常的多核苷酸:ppGpp 和 pppGpp (5' 二磷酸-

鸟嘌呤-3'二磷酸和 5'三磷酸-鸟嘌呤-3'二磷酸)。由于某些历史原因,人们分别把这两个化合物称为魔斑 1 和魔斑 2 (magic spot 1 and 2)。

relA 的产物是一种称作严紧因子 (stringent factor) 的蛋白质,它是负责合成 ppGpp 的一段酶。严紧因子独占性地处在 50S 核糖体上;在正常蛋白质合成期间严紧因子无活性,因此不能合成 ppGpp。只有在以下两种条件下严紧因子才被活化:(1) 50S 亚基必须在完整的 70S 核糖体之中,而且这个核糖体必须已与 mRNA 结合并正参与翻译作用;(2) 核糖体上的 A 位点上必须有一个空载的 tRNA,而不是一个负载着氨基酰的 tRNA。在正常的生长条件下,空载的 tRNA 的浓度比氨基酰 tRNA 的浓度低得多,这就意味着 A 位点只在极少的情况下才被空载的 tRNA 占据。然而,在某种氨基酸缺乏时,正常情况下运输该种氨基酸的 tRNA 就会是空载的。通过这种方式,缺乏氨基酸就导致产生 ppGpp。因此认为 ppGpp 就是使蛋白合成停止的信号;因此也不再产生额外的 rRNA。

为了弄清 ppGpp 是如何调节 rRNA 的合成,人们付出了很大努力。最近获得了 ppGpp 与 RNA 聚合酶结合的某些证据,已经知道 rRNA 和 tRNA 基因的启动子与其他一些操纵子,如 lac、gal 等中的启动子不同,特别是因为 rRNA 基因转录的起始与众不同,它的转录强度是全诱导状态的 lac 操纵子的四倍,这就使人们认识到 rRNA 的启动子是最为有效的启动子。有人提出,当 RNA 聚合酶与 ppGpp 结合时,它就丢失了与 rRNA 启动子结合的能力,但是对于这个很吸引人的假说至今还没有得到证据。

## 沙门氏杆菌鞭毛的相变异

到目前为止所描述的体系都是响应外界物质如糖或氨基酸浓度的变化的细菌操纵子。细菌中的绝大多数生物合成途径与降解途径都在一定程度上采用了类似于上述的操纵子的体系。在这一节中将讨论另一种体系,它是一种不随 RNA 聚合酶能否与启动子结合而决定“开启-关闭”状态的体系。这个体系的开启或关闭是由启动子与结构基因之间的物理联系(或缺乏这种联系)而决定的。

许多细菌都拥有鞭毛(图 14-27),即长长的、像头发似的由蛋白质构成的附件。它能够推动细菌游走于液体介质之间或通过固态的表面,一般是朝向具有所需要的营养的地方,或者是避开有毒物质的地方。鞭毛主要由大量的称为鞭毛蛋白(flagellin)的蛋白亚基构成。

图 14-27 沙门氏杆菌几个细胞的电镜图(承 Mel Simom 氏提供)。

沙门氏杆菌具有两个称为 H1 与 H2 的基因,它们分别负责合成两种不同的鞭毛蛋白。这两种鞭毛蛋白都能聚集形成鞭毛,但它们的免疫性不同。在任何一种特定的细菌中,两个 H 基因中只有一个能转录。因此这个细菌或者形成 H1 型鞭毛,或者形成 H2 型鞭毛。如果 H1 型鞭毛的细菌可以生长许多世代,产生一个很大的细胞群体,那么在这个细菌群体中总会有某些细菌具有 H2 型鞭毛。H2 型鞭毛出现的频率大约是每个世代的每个细胞上有  $10^{-5}$ — $10^{-3}$  个(严格的频率则取决于个别的细菌菌株)。如果筛选出带有 H2 型鞭毛的细菌,并允许它生长,那么在一定的



时间之后,终将显现出一些带有 H1 型鞭毛的细菌,这个频率也是  $10^{-5}$ — $10^{-3}$  (每个世代的每个细胞)。这个现象被称作相变异 (phase variation)。

这两种基因的 DNA 已被分离到,用它们作探针再通过分子杂交实验去检测 mRNA。这些实验的结果表明,在给出的任何一个细菌中只有一种基因被转录。在沙门氏杆菌染色体上 H1 与 H2 基因互相离得很远,这就意味着在相的转变中这种细菌必然使用某些可以扩散的基因产物。我们会看到事实也的确如此。

已分离出 H1 和 H2 基因,并且测出了其中的一部分序列。H1 基因没有什么不平常之处,但是在多次分析 H2 基因时,发现这个基因存在着两种不同的形式: H2 基因的一端是一个 970 碱基对的片段;比较特殊的是,这个片段在一些 H2 基因中是一种方向,而在另一些中却是相反的方向。从 H1 型鞭毛细菌群体中只能分离得到具有一种方向的末端,另一种反向的末端则全部是从有 H2 型鞭毛的细菌中得到的。因此,看来 970 碱基对末端的倒转与 H1 和 H2 基因的活性有关系。了解这种调控的关键就是当这个片段处于一种方向时, H2 被转录;当这个片段的方向相反时, H2 就没有活性,不被转录。图 14-28 给出了对两种活性状态的解释: 可倒转的 970 碱基对的片段中含有 H2 的启动子。因此在这个启动子处,起始一个方向的 mRNA 合成,导致 H2 基因的转录;然而在另一个方向(即“关”的方向),则不对 H2 基因进行转录。因此, H2 的这种倒转说明了 H2 基因的活化与去活化调节,但是未说明 H1 的调控。前面我们曾提到过一个可以扩散的基因产物在这个体系中起调节作用。当 H2 被转录时,形成一个多顺反子 mRNA,这个 mRNA 分子不仅编码 H2 鞭毛蛋白,而且还编码一个抑制 H1 转录的蛋白质。因此,在 H2 顺向时(即定向为有活性)能够形成 H2 型鞭毛蛋白,而不能形成 H1 型鞭毛蛋白,在 H2 反向时(即定向为无活性)被合成出来的是 H1 型的鞭毛蛋白,而不是 H2 型的鞭毛蛋白。

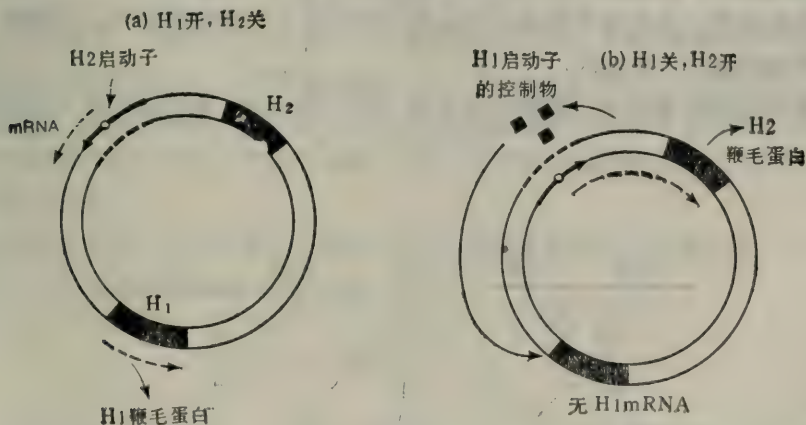


图 14-28 沙门氏杆菌的两种鞭毛蛋白基因的调节。粗虚线表示含有 H2 启动子的可倒转的片段,虚线箭头表示 mRNA 分子。

这个体系是位点专一性倒转 (site-specific inversion) 方式进行调节的一个例子。即 970 碱基对序列从染色体上以某种方式剪切下来,然后以相反的方向再插回去。进一步再对 H2 多顺反子 mRNA 所编码的产物进行分析,结果表明在可倒转的片段中除了有 H2

启动子和编码 H1 转录抑制物的序列之外,还含有一个基因, 它的产物对于倒转作用来说是必需的。

我们还未解释什么物质决定着倒转的发生。在本章前面一些体系的分析中已经指出,某些基因的表达是对一些外部影响作出的反应,虽然目前还未发现影响上述相变异转换频率的外部因素,但是基因表达引起的相变异可能就是属于这种情况。它有可能是一个随机的事件,这在下述的情况中将具有明显的优越性。假设一个动物被一些形成 H1 鞭毛的细菌感染,在细菌生长的同时,动物的免疫系统将产生针对于 H1 鞭毛的抗体,因而这种细菌最终会被动物产生的抗体消灭。然而,如果细菌以较高频率发生转变产生 H2 型鞭毛,这样的细菌就能逃避抗体的杀伤。

相变异的优越性与我们已经讨论过的各种操纵子的优越性有很重要的不同之处。那些操纵子具有打开一个所需要的体系或关闭一个浪费的体系的能力,这对于一个细菌个体来说是很有用处的。然而相变异对于一个细菌个体的生存却并没有什么好处,但是它能保证细菌群体逃过抗体的攻击。细菌的很多体系都很可能沿用了这个原理。

在二十三章我们还将看到把位点专一性交换作用用来程序化地调控真核细胞的基因表达,例如,用在产生抗体的过程中和改变酵母的交配类型的过程中。

## 负 反 馈 抑 制

在这一节将介绍不致变转录的调节作用。某细菌具有脱阻遏作用的 trp 操纵子,如果把这种细菌的培养物突然地加到含有色氨酸的培养基中,那么很快将建立起阻遏作用,进一步只能合成出很少量的 trp 酶系。但是, trp 酶系仍然保留其活性,并且,如果没有第二种调节机理,就会发生色氨酸的浪费性合成以及色氨酸前体和能量的非必要性消耗。通过反馈作用 (feedback) 和末端产物抑制作用 (end product inhibition) 可以避免发生上述的情况。其机理是该途径的末端产物关闭了一些酶的活性。在细菌中,生物合成中的反馈抑制是很常见的。

在合成色氨酸的途径中,前两个反应由 trpD 和 trpE 基因的产物催化 (图 14—29)。

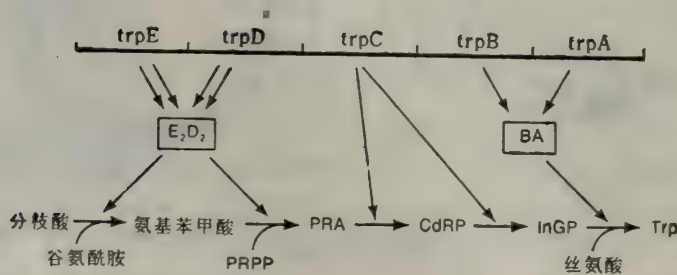


图 14-29 trp 基因产物作用位点。基因 D、基因 E、基因 B 和基因 A 的产物形成有活性的四聚体 (E<sub>2</sub>D<sub>2</sub>) 和二聚体 (BA)。反馈抑制的位置是 E<sub>2</sub>D<sub>2</sub>。这个反应序列的过程中的各个成分缩写为: PRPP: 磷酸核糖焦磷酸; PRA: 磷酸核糖酰氨基苯甲酸; CdRP: 羧基苯氨基-脱氧核酮糖磷酸; InGP: 吲哚甘油磷酸; Trp: 色氨酸。



两种蛋白质的两个亚基聚集在一起组成一个四聚体蛋白质。trpE 亚基催化分枝酸(chorismate)和谷氨酰胺形成氨基苯甲酸(anthranilate)。trpD 亚基再对此产物磷酸化,形成磷酸核糖酰氨基苯甲酸(PRA, phosphoribosyl anthranilate)。trpE 亚基中还含有一个结合色氨酸的位点。但是只有当色氨酸过量时才能结合到它上面。色氨酸占据这个位点会引起 trpE 亚基的构象发生变化。这种变化使酶失去活性(图 14-30) 这个四聚体是一个别构蛋白质(见第六章)。trpE 亚基的构象变化还进一步诱导 trpD 亚基构象发生类似的变化,其活性也因此被抑制。

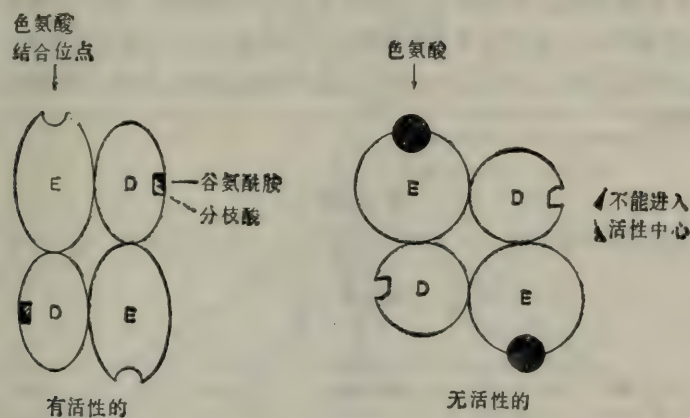
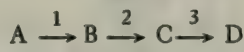


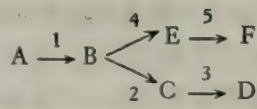
图 14-30 色氨酸使得对形成分枝酸和分解分枝酸起催化作用的 E<sub>2</sub>D<sub>2</sub> 复合体失去活性。

最普遍的反馈抑制类型是生物合成途径的最终产物抑制该生物合成的第一步酶促反应。例如,在以下途径中:



由酶 1、2、3 来催化,产物 D 作用于酶 1。这显然是最经济的一种抑制方式。这就意味着酶 1 必须有两个结合位点。一个是底物结合位点,另一个是产物 D 的结合位点。常常是产物的结合抑制了底物的结合。这是由于两个位点有重叠或是构象发生变化,这种变化减弱或取消了底物的结合位点。在第六章中介绍过第二种机理的例子,即 CTP 抑制了天冬氨酸转氨甲酰酶。

某些生物合成途径是分枝状的。即从一个共同前体能合成出两种产物,下面是这种合成途径的一个假设例子:



在这种合成途径中,不希望有一个产物去抑制酶 1,因为那样会使两个途径都受到抑制。一般来说,最经济的抑制是 D 抑制酶 2、F 抑制酶 4。通过这种方式,D 或 F 都不抑制对方的合成。如果 D 和 F 都存在时,A 向 B 的转换就是一个浪费的过程。因此,在许多途径中发现 D 和 F 都过量时,它们共同去抑制酶 1。这两种不同的分子可以通过四种主要的方式去抑制一个酶促反应步骤:(1) 产物 B 由于酶 2 和 4 被阻塞而积累,酶 1 就被大量的产物 B 所抑制;(2) 酶 1 有被 D 和 F 结合的两个位点,当这两个位点都被占据时,酶就失

活了；(3) D和F对酶1都能分别轻微进行抑制，例如分别抑制d倍与f倍，于是当两个酶共同抑制酶1时，抑制强度就成了 $d \times f$ 倍；(4)在某些途径中，有两种不同的酶（同工酶，isoenzymes）去催化A转化为B，一个同工酶受D抑制，另一个同工酶则受F抑制。

图14-31表示一个分枝状途径的精巧的反馈抑制过程。这是一个研究得比较透彻的多分枝的合成途径，包括从天冬氨酸合成赖氨酸、蛋氨酸、苏氨酸。注意每一种氨基酸如何立刻抑制紧接在分枝后面的那个酶，即酶4、6、8。三个同工酶1a、1b、1c各自被赖氨酸、高丝氨酸和苏氨酸抑制。这三个同工酶具有不同的酶活性；每一个酶都能各自合成一定数量的天冬氨酰磷酸，当这些天冬氨酰磷酸的合成量超过适宜的水平时，它们会反过来分别地抑制相应的同工酶。因此，当一种同工酶受到抑制后，细菌仍然能够合成出正确数量的天冬氨酰磷酸。

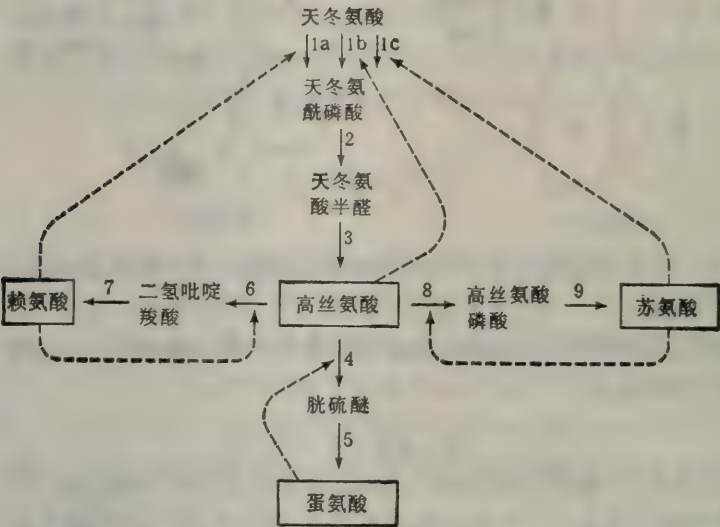


图 14-31 合成赖氨酸、蛋氨酸、苏氨酸途径的反馈抑制作用。框中的分子是抑制物，虚线箭头从抑制子指向被抑制的酶。

### 基因产物对翻译的影响

本章多处提到了翻译的调节，通常这种调节限指以下事实，即多顺反子的 mRNA 翻译出的蛋白质产物拷贝数随不同基因而有所不同。通常，从 mRNA 5' 末端到 3' 末端，翻译能力逐级下降。有几个过程能说明这个现象：不同基因翻译起始效率有所不同，各链的终止密码子与其后面的 AUG 密码子之间的距离不同，mRNA 不同区域对于被降解的敏感性不同。在少数几种细菌噬菌体中发现了一个新的机理，即某个特定基因的翻译过程受到该特定基因的基因产物的抑制。

大肠杆菌噬菌体 R17 仅仅含有三个基因产物，即两个结构蛋白（A 蛋白和噬菌体的外壳蛋白）和一个酶（复制酶）。噬菌体对外壳蛋白的需要量远远大于对复制酶的需要量，复制酶只需有催化量即可。mRNA 分子在外壳蛋白的终止密码和复制酶的 AUG 密码



子之间有一个与核糖体结合位点。由于不断地合成外壳蛋白, 这个结合位点逐渐地被蛋白质分子充满, 核糖体不易再进来, 这就阻抑了核糖体对复制酶区域的翻译作用(图 14-32)。

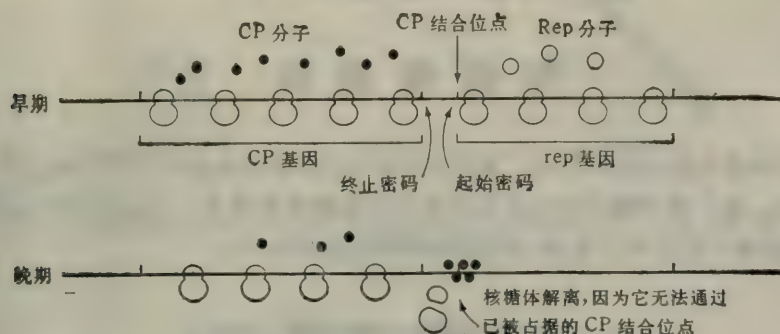


图 14-32 由于外壳蛋白分子的封阻, 噬菌体 R17 的 rep 基因翻译的抑制作用。

## 参 考 文 献

- Clark, B. F. C., H. Klenow, and J. Zeuthen, (eds.). 1978. *Gene Expression*. Pergamon.
- Gilbert, W., and B. Muller-Hill. 1966. "Isolation of the *lac* repressor." *Proc. Nat. Acad. Sci.*, 56, 1891—1898.
- Jacob, F. and J. Monod 1961. "Genetic regulatory mechanisms in the synthesis of proteins." *J. Mol. Biol.*, 3, 318—356.
- Johnson, H. M., W. M. Barnes, F. G. Chumley, L. Bossi, and J. R. Roth. 1980. "Model for regulation of the histidine operon of *Salmonella*," *Proc. Nat. Acad. Sci.*, 77, 508—512.
- Lee, N., W. O. Gielow, and R. G. Wallace. 1981. "Mechanism of *araC* autoregulation and the domain of two overlapping promoters. *pC* and *pBAD* in the L-arabinose regulatory region of *E. coli*." *Proc. Nat. Acad. Sci.*, 78, 752—756.
- Lindahl, L., and J. M. Zengel. 1982. "Expression of ribosomal genes in bacteria." *Adv. Genes*, 21, 53—122.
- Little, J., S. H. Edmiston, L. Z. Pacelli, and D. W. Mount. 1980. "Cleavage of the *E. coli* *lexA* protein by the *recA* protease," *Proc. Nat. Acad. Sci.*, 77, 3225—3229.
- Maniatis, T. and M. Ptashne. 1976. "A DNA operator-repressor system." *Scient. Amer.*, January, pp. 64—76.
- Miller, J. H. 1980. "Genetic analysis of the *lac* repressor." *Curr. Topics Microbiol. Immunol.*, 90, 1—18.
- Miller, J. H., and W. S. Reznikoff (eds.) 1978. *The Operon*. Cold Spring Harbor Laboratory.
- Nierlich, D. P. 1978. "Regulation of bacterial growth, RNA, and protein synthesis." *Ann. Rev. Microbiol.*, 32, 393—432.
- Oxender, D. L., G. Zurawski, and C. Yanofsky. 1979. "Attenuation in the *E. coli* tryptophan operon; role of RNA tertiary secondary structure involving the tryptophan coding region." *Proc. Nat. Acad. Sci.*, 76, 5524—5528.
- Pastan, I., and S. Adhya. 1976. "Cyclic adenosine-3'-5'-monophosphate in *E. coli*." *Bact. Rev.*, 40, 527—551.
- Platt, T. 1978. "Regulation of gene expression in the tryptophan operon in *E. coli*." In Miller, J. H. and W. S. Reznikoff (eds.), *The Operon*, pp. 213—302. Cold Spring Harbor Laboratory.
- Ptashne, M., and W. Gilbert. 1970. "Genetic repression." *Scient. Amer.*, June, pp. 36—44.
- Reznikoff, W., J. Miller, J. Scaife, and J. Beckwith. 1969. "A mechanism for repressor action." *J. Mol. Biol.*, 43, 201—213.
- Reznikoff, W., R. Winter, and B. C. Hurley. 1974. "The location of the repressor-binding site in the *lac* operon." *Proc. Acad. Sci.*, 71, 2314—2318.
- Safer, B., and W. F. Anderson. 1978. "The molecular mechanism of hemoglobin synthesis and its regulation in the reticulocyte." *CRC Revs. Biochem.*, 6, 261—290.
- Ullman, A., and A. Daeschin. 1980. "Role of cyclic AMP in regulatory mechanisms of bacteria." *Trends. Biochem. Sci.*, 5, 95—96.
- Wallace, R. G., N. Lee, and A. V. Fowler. 1980. "The *araC* gene of *E. coli*: transcription and translation start points and complete nucleotide sequence." *Gene*, 12, 179—190.
- Yanofsky, C. 1981. "Attenuation in the control of expression of bacterial operons." *Nature*, 289, 751—758.

## 第十五章 噬菌体 (I)

### 烈性噬菌体

噬菌体对分子生物学的发展起了重要的作用。很多噬菌体是目前了解最清楚、研究最深入的有机体。由于噬菌体的结构比细菌和高等细胞简单得多,而且我们可以获得大量的突变体,噬菌体广泛应用于复制、转录和调节机理的研究中。

#### 噬菌体的一般性质

噬菌体寄生于细菌中。它本身虽然可以生存,但只有寄生时才能够生长和繁殖。尽管多数噬菌体都具有编码各种不同蛋白质的基因,然而目前已知的噬菌体都是利用寄主细胞的核糖体、蛋白质合成因子、氨基酸以及产能装置。这说明噬菌体只能在有代谢活性的细菌中生长。因此,研究噬菌体的生活史必须用感染对数生长期的细菌,以保证细菌具有代谢活性。

噬菌体为维持生存应具有一定的最基本的功能:

1. 保护其本身的核酸免受环境中化学物质的破坏(核酸分子的断裂和突变)。
2. 使其本身的核酸可以侵入细菌内部。
3. 使被感染的细菌产生大量的子代噬菌体。
4. 能使被感染的细菌裂解,释放出噬菌体。本章将详细讨论几种不同种类的噬菌体,从中可以得到以上问题的答案。这些噬菌体具有某些共性,但其细微的差异性却表明特异的生物学功能是通过不同的途径来完成的。

本章应特别引起读者注意的是各种噬菌体对寄主细胞装置的利用程度。基因数少于 10 的噬菌体几乎完全依赖于寄主;基因数在 30 至 50 之间的噬菌体的依赖性就要小些;而一些较大的噬菌体甚至不需要寄主的基因,它们本身含有很多基因,可独立完成 DNA 复制等功能。令人吃惊的一点是,某些噬菌体的基因几乎就是寄主基因的翻版,这是功能上的重复吗?我们将在后面的“非必需基因”的题目下讨论。

噬菌体因种类不同而具有不同的物理结构,因而也就有不同的生活史。后面我们也将涉及到其结构上的区别。

读者最好能先复习一下第一章中的噬菌体生物学及噬菌体的检出和计数技术。

#### 噬菌体的结构

噬菌体有三种基本结构:20 面体无尾部噬菌体、20 面体有尾部噬菌体和丝状噬菌体。噬菌体一般只含有一种核酸分子(单链的或双链的线性 DNA,环状 DNA,或单链线性 RNA);一种或几种蛋白质分子(已知的唯一例外是  $\phi 6$  噬菌体,它有所谓的局部基因组:每个  $\phi 6$  噬菌体含有碱基序列不同的三个线性双链 DNA 分子)。蛋白质构



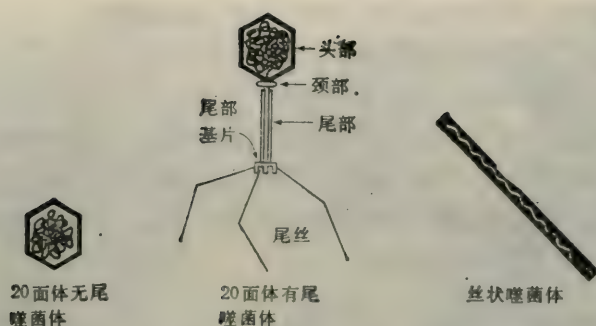


图 15-1 三种噬菌体结构的二维图谱。有尾噬菌体并不总是含有颈部，并可以根据不同种类有着不同数量的尾丝：0—6 根尾丝。

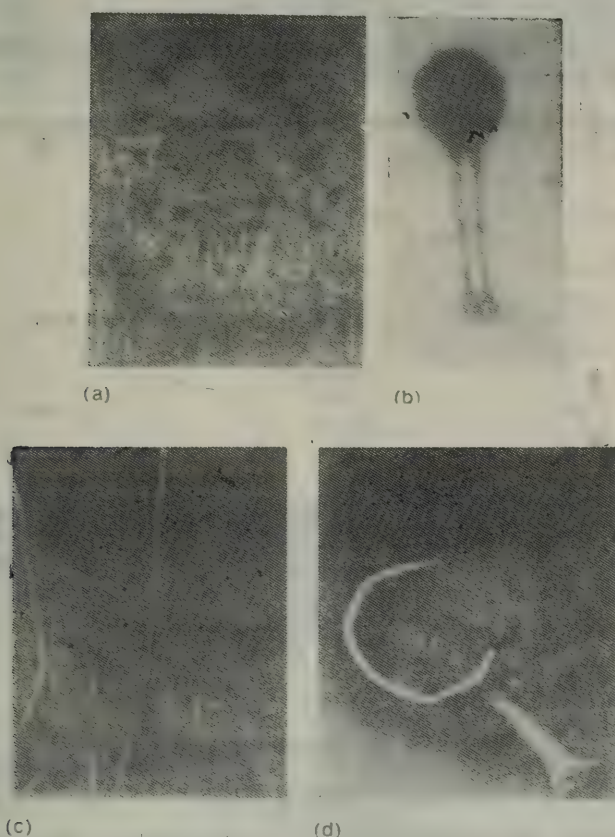


图 15-2 噬菌体的三种主要形态。(a) 20 面体无尾的：PM-2；(b) 20 面体有尾的：SPO1；(c) 丝状体的：M13(承 Edward Kellenberger 提供)；(d) 从其破裂的头部释放出 DNA 的：T4 噬菌体(承 Robley Williams 提供)。

成的外壳——核衣壳包住了核酸，使之不被核酸酶水解，不受有害物质侵蚀。图 15-1 和图 15-2 是这三种基本结构的模式图。要点如下：

1. 在无尾部和有尾部的 20 面体噬菌体中，核酸紧密缠绕而存在于外壳的中空部位。在丝状噬菌体中，伸展的螺旋状的核酸被包埋在外壳里。

2. 有尾噬菌体中,有一段很小的 DNA 从外壳伸向尾部。
3. 尾部是终止于尾丝的复合结构。
4. DNA 分子比其头部任何角度的长度大得多。

有尾噬菌体的结构有很多变化。如头部的长度可大于或等于其宽度,小于其宽度的还未见到。最短的尾部仅仅在电镜下可见,而最长的可达其头部长度的 4 倍。尾部的基板可能是非常复杂的,但也可能根本就不存在;它存在时,通常有 1—6 根尾丝。另外,围绕头尾连接点的颈部也并非一定存在。

### 典型噬菌体的裂解途径

噬菌体的生活史分为两类——裂解和溶源。裂解反应中,噬菌体可使被感染的细胞释放出很多子代噬菌体。仅能裂解生长的噬菌体叫做烈性噬菌体 (virulent)。溶源反应只在双链 DNA 的噬菌体中观察到,这类噬菌体称为温和噬菌体 (temperate)。侵入寄主后,噬菌体的 DNA 整合为细菌染色体的一部分,并不产生子代噬菌体。本章只介绍裂解途径,在研究几种特殊种类的烈性噬菌体后,下一章再来讨论溶源反应。

不同的烈性噬菌体的生活史各不相同。我们先介绍一种基本的裂解途径(图15-3):

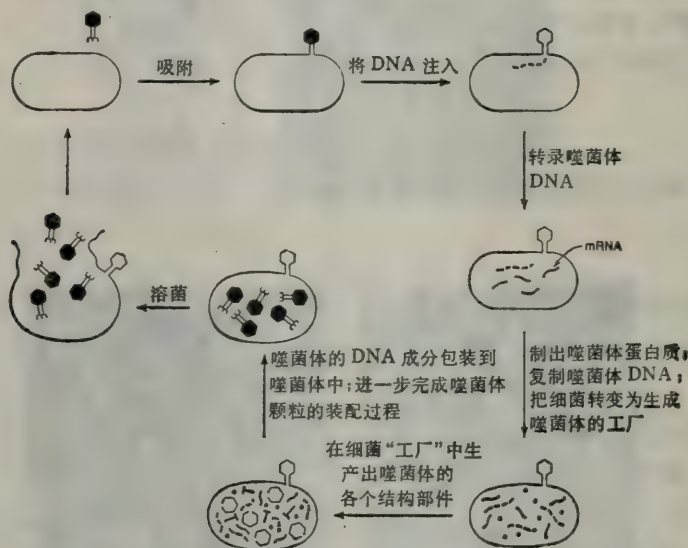


图 15-3 一个典型的噬菌体生活周期图解。

1. 噬菌体对细菌表面专一性受体的吸附(图 15-4)。这些受体形态差异很大,它们是细菌本身服务而不是为噬菌体吸附服务的(因为噬菌体吸附对细菌来说无疑是一个致死因素),例如某些是用于转运蔗糖的蛋白质(大肠杆菌  $\lambda$  噬菌体),某些用于运输  $\text{Fe}^{2+}$  的蛋白质(大肠杆菌 T1 噬菌体)。

2. 噬菌体 DNA 通过细菌细胞壁进入细菌细胞的通道。图 15-5 介绍了有尾噬菌体 DNA 注射进入细胞的过程。这里,核酸不暴露在受体细胞周围的介质中而是直接转运到细胞中。其他几种类型噬菌体的转运机理还不很清楚。

无尾噬菌体的核酸对核酸酶瞬时性地敏感,因此认为可能是噬菌体外壳先破裂,释放其核酸于细胞壁上,然后再进入细胞。





图 15-4 许多噬菌体颗粒通过尾部吸附在大肠杆菌的电子显微镜照片。吸附的噬菌体是  $\lambda$ ，而不是 T4，但是在吸附时这两种类型的噬菌体的外观是非常相似的。

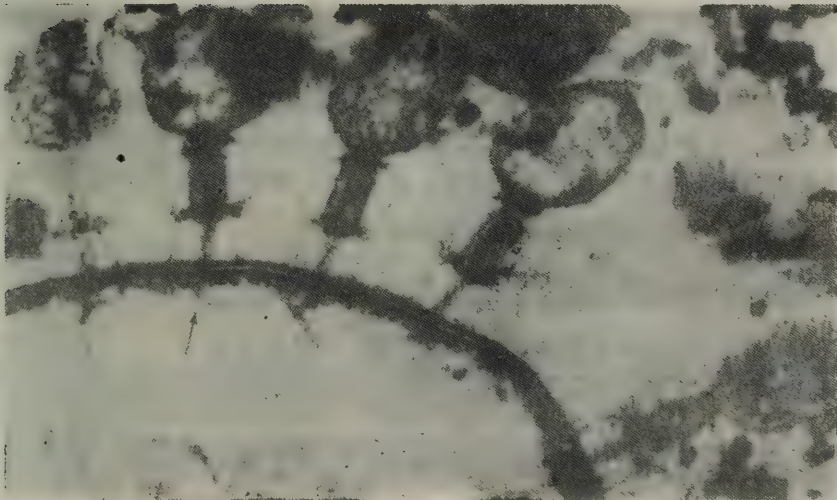
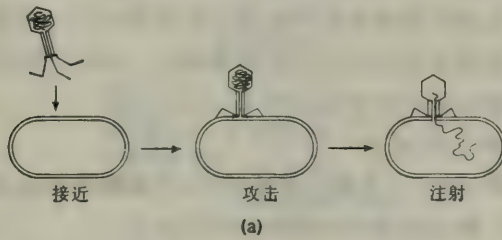


图 15-5 (a) 有尾噬菌体的注射顺序过程。注射时尾鞘卷缩，颈髓部蛋白质插到细菌的细胞壁中，像进行皮下注射那样将 DNA 注入。(b) T4 噬菌体吸附于大肠杆菌细胞壁时的电镜照片。尾鞘卷缩，颈髓部蛋白质坚固地固定到细菌的细胞壁中。箭头所指的是髓部经过细胞壁而射出的部分。从右边的两个噬菌体可以看到 DNA 进入到细胞中。

3. 被感染的细菌转变为产生噬菌体的细胞。被噬菌体 DNA 感染后,很多细菌丧失了复制或转录本身 DNA 的能力,或同时丧失这两种功能。寄主 DNA 或 RNA 合成的关闭过程因噬菌体的种类而异。

4. 噬菌体核酸和蛋白质的产生。噬菌体通过合成噬菌体专一性聚合酶或在细菌的酶中加入某些特殊因子,使细菌细胞复制系统专一性地复制噬菌体核酸。上述两种情形都使用了许多细菌的复制蛋白,因而 RNA 噬菌体局限性很大——细菌不含复制 RNA 的酶,它必须编码本身的复制酶。

转录一般都由细菌的 RNA 聚合酶起始,后面的过程则由以下两种酶之一来催化:被修饰后能识别噬菌体启动子的细菌聚合酶或噬菌体的专一的 RNA 聚合酶。

转录是被调控的,噬菌体的蛋白质在需要时及时有序地合成。

RNA 噬菌体所含的单链 RNA 作为其本身的 mRNA,不需转录,这些噬菌体通过控制其翻译效率而受到调控。

本章后半部分将介绍几种调节过程。

一般来说合成的核酸和蛋白质的数量可以组装 50—100 个噬菌体,具体数量由噬菌体的种类而定。

5. 噬菌体颗粒的装配。这个过程又称为形态发生过程。装配过程需要两类蛋白质:结构蛋白和催化蛋白。前者以被修饰的形式存在于噬菌体中;后者参与装配过程但并不出现在噬菌体颗粒中。催化蛋白还包括一类成熟蛋白,可将细胞内的噬菌体 DNA 转化为适于被包装进入  $\lambda$  噬菌体颗粒的形式。

20 面体噬菌体的包装过程如下:

(1) 积累噬菌体结构蛋白,形成噬菌体的头部;有尾噬菌体还形成尾部;(2) 核酸浓缩入组装好的头部;(3) 尾部附着到充满核酸的头部上。

丝状噬菌体的核酸和蛋白质的形成只需一步,浓缩的机理尚不清楚。

6. 新合成噬菌体的释放。多数噬菌体在感染的后期合成一种蛋白质:溶菌酶或内溶菌素(endolysin),可使细菌细胞壁破裂;或合成另一种蛋白质:膜解酶(membranases),溶解细胞膜。在这两种酶的作用下,细胞被完全毁坏,噬菌体释放到周围介质中,此过程叫做裂解(lysis)。几种丝状噬菌体通过细胞壁的突起连续地释放子代噬菌体——出芽过程。被这类噬菌体感染的细胞并不被降解,而能在一段较长的时间内连续产生病毒颗粒。

### 噬菌体感染的细菌培养物的性质

前面我们描述了细胞被噬菌体感染后的历程。实验室条件下,一般都使用大量噬菌体颗粒去感染菌液。下面介绍几个参数,在本节和下面三节中我们将使用它们来描述此种过程。

噬菌体对菌落的吸附是一个随机的、无序的过程,泊松定律可用于表达这种随机吸附时细菌中噬菌体分布的情况:

$$p(n) = m^n e^{-m} / n!$$

在这里,当被吸附的噬菌体数为  $n$ ,每个细菌平均所吸附的噬菌体数(即感染倍数 multiplicity of infection,或简称为 MOI)为  $m$  时, $p(n)$  为被感染的细菌的比例数。



即用  $0, 1, 2, 3, \dots, i$  个噬菌体感染时,所感染的细菌的比例数为  $p(0), p(1), \dots, p(i)$ 。因此,如果有  $3 \times 10^8$  个噬菌体吸附在  $10^8$  个细菌上时 ( $m = 3$ ), 则  $p(0), p(1), p(2), p(3) \dots$  为  $0.05, 0.15, 0.22, 0.22, \dots$ 。由于  $p(0) = 0.05$ , 所以  $p(1) + p(2) + \dots + p(i)$  就等于  $1 - 0.05 = 0.95$ 。也就是说,有 95% 的细菌被至少 1 个噬菌体感染。

另外,  $p(0)$  的值亦可说明已吸附的噬菌体的比例。例如,在以上的感染过程中,假若所有加入的噬菌体都吸附在细菌上,对于  $m = 3$  来说,  $p(0)$  就应等于 0.05; 然而在某一实验中使用了  $3 \times 10^8$  个噬菌体,  $10^8$  个细菌,却观察到有 12% 的细菌未被感染,  $p(0)$  等于 0.12, 运用泊松定律  $p(0) = e^{-m}$  演算,将得出  $m = 2.12$ ——MOI 的真实值  $2.12/3 = 0.71$ 。因此加入的噬菌体实际上只有 71% 吸附到了细菌上面。

$p(0)$  可以用简单的方法测量(图 15-6)。将经过吸附期的细菌涂布于琼脂表面,我们再测量可以形成菌落的细菌的比例数——只有未被感染的细菌才可以形成菌落。

被感染的细胞可用如下方法测定。首先,将能使未被吸附的噬菌体失活的抗体\*\*加入感染了的培养物中,再把噬菌体悬浮液倒在对噬菌体敏感的细菌所形成的菌苔上,此细胞由于是在裂解前涂布于平板上的,每个被感染的细胞将只产生一个噬菌斑。

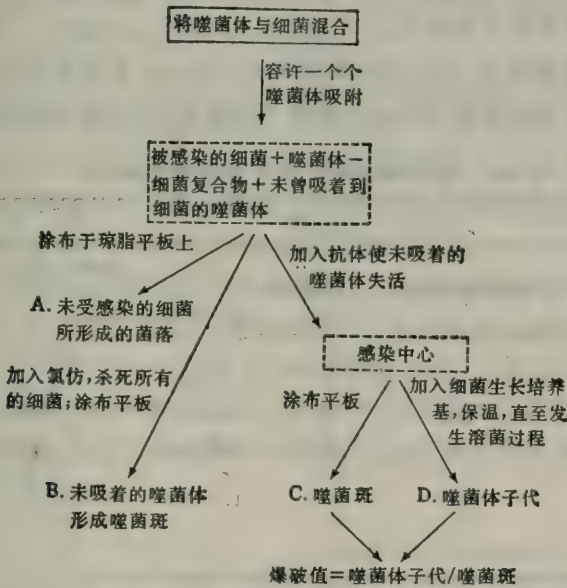


图 15-6 测定噬菌体感染实验值过程的图解。A. 测定未感染的细菌数。B. 测定未吸着的噬菌体数。C. 测定感染中心。D. 测定子代噬菌体数及爆发值 (D/C)。

通过此种途径形成一个噬菌斑的一个细胞称为一个感染中心(infective center)。在吸附期之后,通过加氯仿至培养液中,可测出未被吸附的噬菌体数目:游离的噬菌体不受氯仿处理的影响,而感染中心则将被氯仿杀死。因此,加入氯仿后形成的噬菌斑数就代表未吸附的噬菌体数。

在进行得很好的实验中,常常要测量感染中心的数目、游离噬菌体的数目以及 MOI

\* 原书误为  $1-0.005$ 。——译者注  
\*\* 抗体来自注射了纯噬菌斑悬浮液的兔子的血液。

的真实值,因为许多实验结果都依赖于 MOI 值。

一个被感染细胞所产生的噬菌体数叫做噬菌体爆破值 (burst size)。在许多实验中,它都是一个重要的参数,因为它是噬菌体产生效率的量度。将细菌培养物裂解后产生的噬菌体数(通过对噬菌斑进行计数而得到)除以感染中心的数就可得到此值(见图15-6)。

噬菌体的增殖比细菌快得多。细菌经过一个世代才增加一倍,而噬菌体经过一个生活周期后的增加量却等于它的裂解量,表 15-1 可明显地看到这一点。这里,一个平均裂解量为 100,生活周期为 25 分钟的噬菌体感染 1 毫升增殖周期为 25 分钟的细菌培养物。计算时,假定吸附是迅速而彻底的。请注意,四个世代以后,细菌数量增加了 14 倍,而噬菌体数值则增加了  $10^8$  倍,此时噬菌体数约为细菌数的 8 倍,因此,所有的细菌都会被感染。所以,在 125 分钟后,细菌全都裂解了,而原来的噬菌体则产生了  $1.4 \times 10^9$  个子代噬菌体。

噬菌体的必需和非必需基因

本章其余篇幅主要用于讨论各种噬菌体。对于它们,什么是必需和非必需基因呢?如果噬菌体的一个基因突变后仍能形成噬菌斑,这个基因就是非必需基因;反之,突变后不能形成噬菌斑的基因就是必需基因。对这个定义应该小心地说明,因为对于一个被感染细菌来说,典型的爆破值为 50—100,爆破值为 4 就足以形成噬菌斑。实际上,许多所谓的非必需基因的突变尽管还不足以妨碍形成噬菌斑,但仍能显著地减小爆破值。

表 15-1 若干细菌世代后噬菌体和细菌数值的计算

世 代 数	浓 度		
	噬菌体数/毫升	细菌数/毫升	
		近 似 值	精 确 值
0	1	$10^6$	$10^6$
1	$10^2$	$2 \times 10^6$	$2 \times (10^6 - 1)$
2	$10^4$	$4 \times 10^6$	$(4 \times 10^6) - (2 \times 10^2) - 4$
3	$10^6$	$7.98 \times 10^6$	$(8 \times 10^6) - (2 \times 10^4) - (4 \times 10^2) - 8$
4	$10^8$	$1.4 \times 10^7$	$(1.6 \times 10^7) - (2 \times 10^6) - (4 \times 10^4) - (8 \times 10^2) - 16$
5	$1.4 \times 10^9 = 1.4 \times 10^7 \times 100$	0	0

\* 细菌培养起始时用 1 个噬菌体感染  $10^6$  个细胞/毫升浓度的细菌菌液。细菌的增殖周期与噬菌体的生活周期时间相同。

因为下述四种原因,基因可能是非必需基因:

(1) 细菌中有与之结构相同或功能相同的基因。此种加倍可以提高必需酶浓度而使噬菌体爆破值提高,因而可能对噬菌体是有利的,另一方面,自然界中有些宿主可能缺乏某个基因,而它对于噬菌体生长来说,却是必需的;(2) 这个基因并不与细菌基因完全相同,而仅仅通过某种途径可提高噬菌体产生的速度或爆破值。上述两点在进化上都是有利的;(3) 实验室条件下并不需要这个基因,但在自然界中需要应付一些特殊情况时需要它去倍增噬菌体;(4) 另外,还有一种情况,即经常地不需要这个基因。这种可能性很小,因为噬菌体不会保留一个真正无用的基因。



我们将看到所谓的非必需基因仅仅存在于中等的和大的噬菌体中。

### 噬菌体感染的专一性：吸附作用

现已分离出几千种噬菌体。一种噬菌体只能感染一个种属，而且常常只是一个种属里的几株细菌\*。例如，没有一种既能感染假单胞菌属 (*Pseudomonas*) 又能感染大肠杆菌的噬菌体；甚至能在 *Ps. fluorescens* 中生长的噬菌体，不能生长于 *Ps. aeruginos* 中。在研究得最广泛的大肠杆菌的噬菌体中，可以观察到这种高度特异性。如  $\phi$ X174 可在大肠杆菌 C 株中很好地成活，而在大肠杆菌的其他实验菌株内却不能成活。不过，也有例外，如大肠杆菌的 T4 噬菌体能在大肠杆菌的很多菌株和志贺菌 (*Shigella*) 属中的一些种中生长。

这种专一性是由好几种因素决定的。其一为吸附能力。细菌细胞壁上存在有受噬菌体吸附的专一受体，所以大肠杆菌的噬菌体不能吸附到任何假单胞菌属上。噬菌体吸附在细菌上，并杀死细菌，对细菌来说无疑是一件很不好的事，因而在进化时，细菌细胞壁的成分要发生变化，使大多数噬菌体不能识别它；同时，噬菌体也在演化，通过这样地变化去识别一些细菌。如果某一种噬菌体不会这样地去识别细菌，它也就不能生存。实验室中的一个小实验可以为这种进化提供一个例证。

用  $10^{10}$  个 T6 噬菌体感染  $10^8$  个大肠杆菌 B 细胞，被感染的细胞接种在琼脂表面，约形成 100 个菌落。由这些菌落繁殖的细菌，丧失了被 T6 感染的能力，噬菌体不能再吸附于突变的细胞上(当然这些细胞不是由非感染过程繁殖产生，而是感染后存在于培养物中的；并通过与过量噬菌体生长在一起而筛选到的)。这些细菌的突变株叫做 T6 抗性株(图 15-7)，用 B/6 或  $T_{sx-r}$  表示。若琼脂培养基上有  $10^8$  个  $T_{sx-r}$  细胞形成的菌苔，再加入  $10^8$  个 T6 噬菌体，约形成 10 个噬菌斑。噬菌斑中的这些噬菌体是携带着尾部纤维基因突变体的，有吸附于  $T_{sx-r}$  细胞的能力。这类噬菌体叫做 h 突变体(寄主范围)。通常它们还有可以在  $T_{sx-s}$  细菌上形成噬菌斑的能力，被认为有较大寄主范围。

在几种情况下，吸附是跨种感染的唯一障碍。T4 噬菌体不能吸附于产气杆菌 (*Aerobacter*) 属便是一个例子，但使用某些特殊的技术可迫使 T4DNA 进入 *A. aerogenes* 的原生质体，产生子代噬菌体。

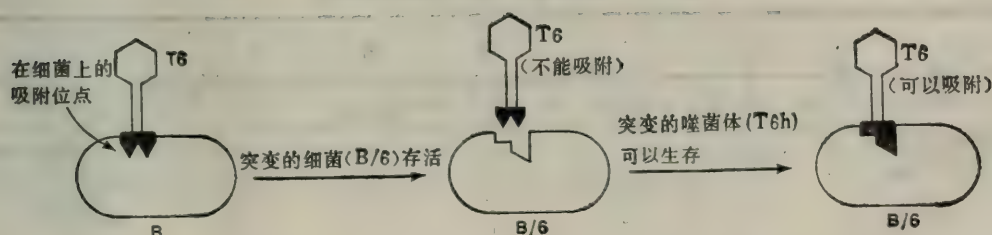


图 15-7 抗噬菌体的抗性细菌以及噬菌体的 h 突变体的产生。尽管突变时只是尾丝受到修饰，但为了图解的需要，尾板也被修饰。

\* 一个有趣的例外是 PRR<sub>1</sub> 噬菌体，它能感染携带 RP4 质粒的任何革兰氏阴性菌。

噬菌体感染的专一性：寄主限制

前面我们谈到了噬菌体不具有对大多数细菌专一吸附的能力，但对某些细菌来说更重要的一个障碍是“外源”的细菌 RNA 聚合酶不能识别噬菌体启动子。甚至当噬菌体可以进行吸附和转录时，多数细菌还有宿主限制和宿主修饰的问题。也就是说 X 型细菌可以区分在 X 型细菌中生长的和在其他如 Y 型细菌中生长的噬菌体，并能防止在 Y 上生长的噬菌体对 X 进行有较的感染。这里所涉及的重要符号为：在 X 型细菌中生长的噬菌体 P 记为 P · X。宿主限制和宿主修饰的资料见表 15-2。注意，在大肠杆菌 K 菌株上正常而在 B 株上形成效率极低的噬菌斑的  $\lambda \cdot K$  是被 B 菌株限制的。在所形成的极少量的噬菌斑 ( $\lambda \cdot B$ ) 中的噬菌体已经被 B 株修饰，可以在 B 上有效生长，然而， $\lambda \cdot B$  却不能在 K 株上生长了——它被 K 株限制了。从分子水平来说，就是大肠杆菌 B 含一种限制性内切酶或一种限制酶——在这里就是 EcoB 核酸酶，它是一种位点专一的核酸酶，只能切割 DNA 链的特异碱基序列。 $\lambda \cdot K$  就含有这个序列；当  $\lambda \cdot K$  的 DNA 射入大肠杆菌 B 时，它就被切断。大肠杆菌 B 也含这种序列，若未被修饰，同样也会被水解；一种位点专一的甲基化酶 (EcoB 甲基化酶) 对这段序列中的腺嘌呤甲基化，使得这段序列不被 EcoB 核酸酶水解。当  $\lambda \cdot K$  感染 B 株时，大量被感染细胞内的几个亲代噬菌体 DNA 分子在限制前被甲基化，因而产生少量 B 型修饰 ( $\lambda \cdot B$ ) 的噬菌体。大肠杆菌 K 同样也含有一种位置专一性的核酸酶叫做 EcoK 酶，它与 EcoB 识别的序列不同，一种 EcoK 甲基化酶也能产生 K 修饰，使大肠杆菌 K 的 DNA 不自我降解。能在 K 株中生长的  $\lambda$  噬菌体  $\lambda \cdot K$ ，在 EcoK 的专一性序列甲基化，从而对 EcoK 酶产生抗性。 $\lambda \cdot B$  有未甲基化的 EcoK 序列，当  $\lambda \cdot B$  DNA 感染 K 株细胞时总是被降解。 $\lambda \cdot B$  DNA 分子偶尔因复制而未被限制，其复制物则含有一个甲基化的 K 专一性序列。因此，当  $\lambda \cdot B$  有效感染大肠杆菌 K 时，产生少量的子代噬菌体为  $\lambda \cdot K$ ，它不再有 B 修饰了，当它感染 B 时就会被限制。

表 15-2 中，C 株中生长的噬菌体  $\lambda \cdot C$ ，不能在 B 和 K 株中生长，C 株对  $\lambda \cdot B$  和  $\lambda \cdot K$  却都不限制，这是由于 C 株不含水解  $\lambda$  DNA 内任何碱基序列的限制性内切酶。C 株则不含 B 和 K 甲基化酶，B 和 K 菌都将限制  $\lambda \cdot C$  噬菌体。

表 15-2 大肠杆菌  $\lambda$  噬菌体的限制和修饰

菌 株	噬 菌 体		
	$\lambda \cdot K$	$\lambda \cdot B$	$\lambda \cdot C$
K	1	$10^{-4}$	$10^{-4}$
B	$10^{-4}$	1	$10^{-4}$
C	1	1	1

注意：数字表示相对的涂布效率。

宿主限制和修饰是一个普遍过程，它也许是用于破坏外源 DNA 的。在第二十章中将可以知道，限制性内切酶是重组 DNA 技术的基础。



# 噬菌体专论

从生物学水平上看,噬菌体是一种相对简单的生命形式。实际上它也十分复杂,还没有一种噬菌体的生命过程的分子机理被完全搞清楚。主要的进展是通过研究生长在大肠杆菌、枯草杆菌 (*Bacillus subtilis*) 和沙门氏杆菌 (*Salmonella typhimurium*) 的几种噬菌体而得到的。不同的噬菌体用于研究不同的过程可能收效更大: 用  $\lambda$  和 T7 噬菌体得以最清楚地了解转录的调控; 用 T4 和  $\lambda$  噬菌体研究形态发生取得很大成功; P1 和 P22 噬菌体的研究首先提供了转导的材料; 用 T5 说明 DNA 的侵染; T4 用于研究 DNA 的合成; T4 和 T7 用于研究噬菌体如何取代细菌等,都取得重要进展。

下面我们将详细介绍其中的几个(及另外几个)噬菌体。重点放在了解得很清楚的性质上,其他性质仅简述。尽可能介绍每个噬菌体的生活周期。表 15-3 列举了几类噬菌体的某些重要性质。

表 15-3 几种噬菌体核酸的性质

噬菌体	宿主	DNA 或 RNA	形式	分子量 $\times 10^4$	稀有碱基
$\phi$ X174	E	DNA	ss,circ	1.9	无
M13,fd,f1	E	DNA	ss,lin	2.1	无
PM2	PB	DNA	ds,circ	9	无
186	E	DNA	ds,lin	18	无
B3	PA	DNA	ds,lin	20	无
Mu	E	DNA	ds,lin	25	无
T7	E	DNA	ds,lin	26	无
$\lambda$	E	DNA	ds,lin	31	无
N4	E	DNA	ds,lin	40	无
P1	E	DNA	ds,lin	59	无
T5	E	DNA	ds,lin	75	无
SPO1	B	DNA	ds,lin	100	HMU 代 T
T2,T4,T6	E	DNA	ds,lin	108	葡萄糖基化 HMC 代 C
PBS1	B	DAN	ds,lin	200	U 代 T
MS2,Q $\beta$ ,f2	E	RNA	ds,lin	1.0	无
$\phi$ 6	PP	RNA	ds, lin	2.3	无
				3.1	
				5.0*	

注意: 表中使用的缩写含意如下: E: 大肠杆菌; B: *Bacillus Subtilis*, 枯草杆菌; PA: *Pseudomonas aeruginosa* 假单胞菌; PB: *Ps. aeruginosa* BAL-31, 假单胞菌 BAL-31; PP: *Ps. phaseolica*, 假芸豆菌; SS: single strand, 单链; ds: double strand, 双链; circ: circular, 环状; lin: linear, 线状; HMU: hydroxymethyluracil, 羟甲基尿嘧啶; HMC: hydroxymethyl cytosine, 羟甲基胸嘧啶。

\* $\phi$ 6 含有三个分子。

## 大肠杆菌 T4 噬菌体

1939—1941 年间, Milislav Demerec 分离出一系列生长于大肠杆菌的 B 菌株上的噬菌体, T1, T2, ... T7。凭借这些噬菌体, 把生化实验和遗传学分析联系起来, 使遗传学发生了一场革命。后来发现这些噬菌体的遗传材料为双链 DNA (第七章), 它们的结

构和生活史有很大区别,据此把它们分为 4 组: T1;T2、T4、T6 (T 偶数噬菌体); T3、T7; 和 T5\*。

T 偶数噬菌体容易大量生长,也易获得很多突变株,它的 DNA 分子含有一个稀有碱基使之与宿主 DNA 相区别。因此, T 偶数噬菌体成为热门的研究对象。近年来的主要工作是使用这一组中的 T4 噬菌体,其电镜照片见图 15-8。尽管其分子较大,生活周期较复杂仍有一些未知数,我们还是获得了大量知识。其他类型的只有 T5 和 T7 才有较详细的分子信息;我们将在分析完 T4 的生活周期后讨论 T7 的性质。

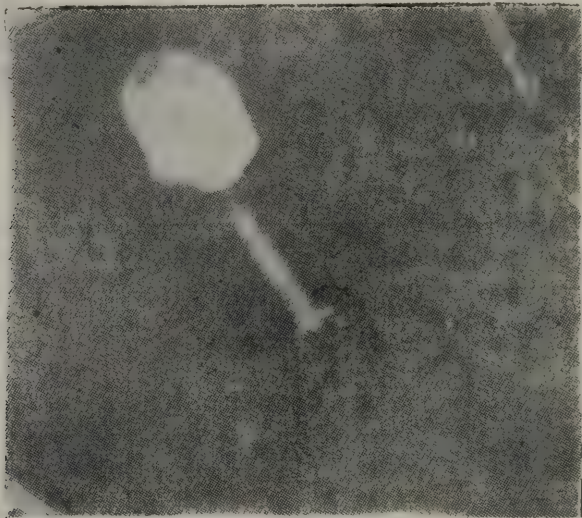


图 15-8 大肠杆菌 T4 噬菌体。噬菌体头部含有 DNA,尾端部的极板上伸出尾丝(承 Robley Williams 提供)。

T4 DNA 的性质

T4 作为一个噬菌体来说,其 DNA 相当大,分子量约为  $1.1 \times 10^8$ , 含 166 000 个碱基对,长为 55 微米,头部宽约 0.06 微米,长约 0.09 微米。当 DNA 被包装时,它需紧密

表 15-4 在 T 偶数噬菌体 DNA 中的 5-羟甲基胞嘧啶 (HMC) 的葡萄糖糖基化作用方式

葡萄糖糖基化作用	总 HMC 的%		
	T2	T4	T6
非糖基化的	25	0	25
$\alpha$ -糖基化	70	70	3
$\beta$ -糖基化	0	30	0
$\alpha$ -糖基化- $\beta$ -糖基化	5	0	72

\* T1、T3、T5 和 T7 又称 T 奇数噬菌体,但因为它们的分属三类,这个名称意义不大。



盘绕 (参见图 15-2)。其 DNA 含 A、T、G，不含 C，C 由被修饰的胞嘧啶，即 5-羟甲基胞嘧啶 (HMC) 代替，HMC 与 G 配对 (见图 15-9)，HMC 通过糖苷化而进一步被修饰——一种糖与 HMC 上的羟基相连。上面提到的三种十分相似的 T 偶数噬菌体只是在糖的成分上有所不同。HMG 上有足够多的糖将 T4 DNA 包围起来，因此纯 T4 DNA 对各种 DNase 有抗性。

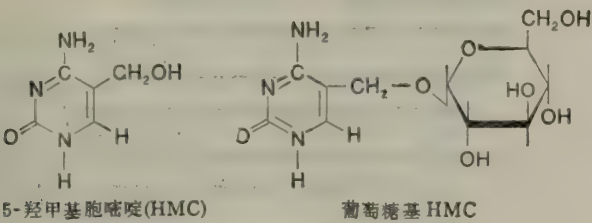


图 15-9 非葡萄糖基化及糖基化的 5-羟甲基胞嘧啶。如果 HMC 中的 CH<sub>2</sub>OH 是 H，那么这个分子就是胞嘧啶。

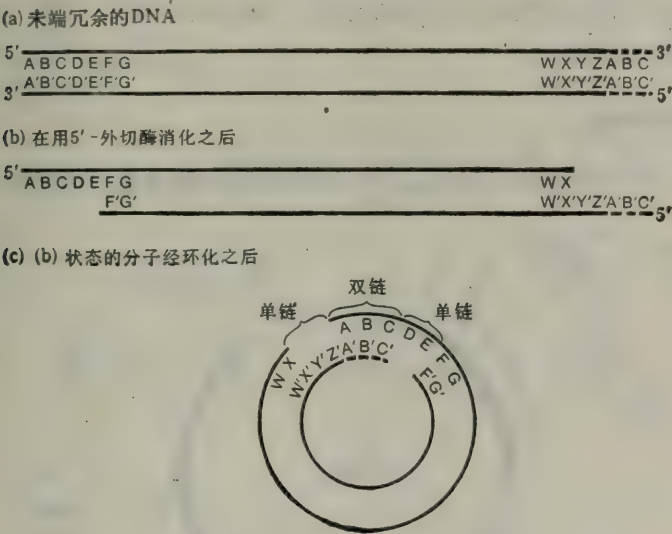


图 15-10 末端冗余分子以及通过外切酶消化和环化的鉴定。非冗余的 DNA 不能通过本法环化。

T4 DNA 的碱基顺序在分子的两端是重复的，这种现象叫做末端冗余 [图 15-10(a)]。在多种噬菌体中出现的末端冗余现象是这样证明的 [图 15-10(b,c)]：将 DNA 样品用 DNA 外切酶处理，从每条单链的 3'-OH 末端依次切去核苷酸，DNA 样品不再能复性，一旦被移去的碱基数目超过末端冗余区域的碱基数时，就可形成一环状结构，它含有相当于末端冗余长度的一段短的双链结构，研究发现末端冗余长度为 1 600 碱基对。

T4 DNA 还有一个重要的性质：虽然每个噬菌体仅含 1 个 DNA 分子，但每个噬菌体的 DNA 分子各不相同，甚至由一个噬菌体和其子代所产生的群体的 DNA 分子也不

尽相同。T4 DNA 分子是环状排列的，见图 15-11。此图系统地说明了在种群中 DNA 分子的末端碱基可以是整个序列中的许多不同碱基，确切地说，也许是整个序列中的任何一个碱基。请注意，环状排列是噬菌体种群的性质，而末端冗余则是个体噬菌体 DNA 分子的性质。

图 15-11 末端重复 DNA 分子的环状排列。

### T4 噬菌体的基因构造

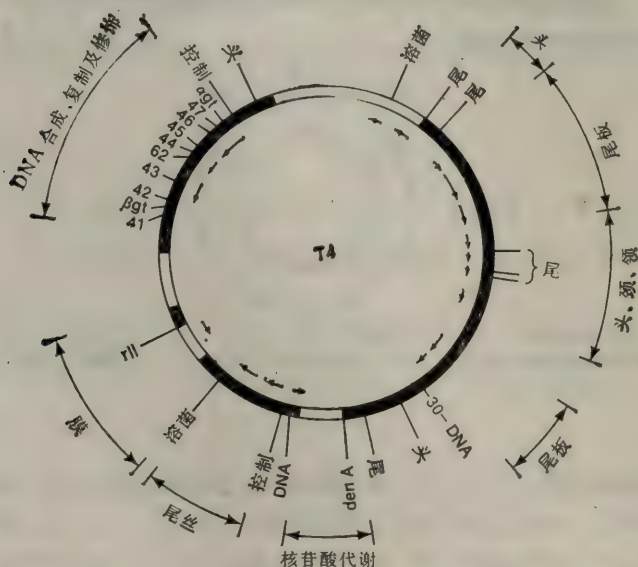


图 15-12 T4 噬菌体的遗传学图谱——表示了部分的但不是全部的基因。从图谱中可以见到具有相关功能的一些基因聚集成簇。虽然那些编码尾板及尾丝基因聚成一大簇，但是还有一些形成尾部其他部分的基因散开分布于图谱中。图中黑色区表示必须基因区，空白区表示非必须基因区。“控制”表示那些起始各种的转录作用所需要的基因。图谱圆圈内的箭头表示各种转录作用的方向及起点(但不是表示长度)。



还有 15—20 个基因尚待发现。基因的遗传学图谱见图 15-12。功能相关的基因常常相邻,并可作为多顺反子 mRNA 分子的一部分而转录。这种排列是非常有效的,使得功能相关的蛋白质的合成几乎同时进行,并可减少所需的调节元素的量。但是并非所有功能相关的基因都作为一个转录单位,某些转录单位也含有与功能无关的基因。将功能相关的基因聚集在一起是很多噬菌体系统的普遍现象,几乎无一例外。后面我们可以看到大肠杆菌  $\lambda$  噬菌体的这种性质。

T4 基因可分为两类: 82 个代谢基因和 53 个颗粒装配基因。82 个代谢基因中仅有 22 个是必需的,它们参与 DNA 合成,转录,裂解;其他 60 个代谢基因只是细菌基因的翻版,这些基因发生突变的噬菌体仍可以生长,只是有时裂解量较少。在 53 个装配基因中,34 个编码结构蛋白,19 个编码催化装配所需的酶和蛋白质因子。因此, T4 DNA 基因的 17% 编码与主要代谢过程相关的蛋白质,39% 用于噬菌体的装配,44% 编码与次要代谢过程有关的蛋白质。

将 T4 基因功能分类的很有用方法见图 15-13,待谈到生命活动的各步时,这些功能就很清楚了。

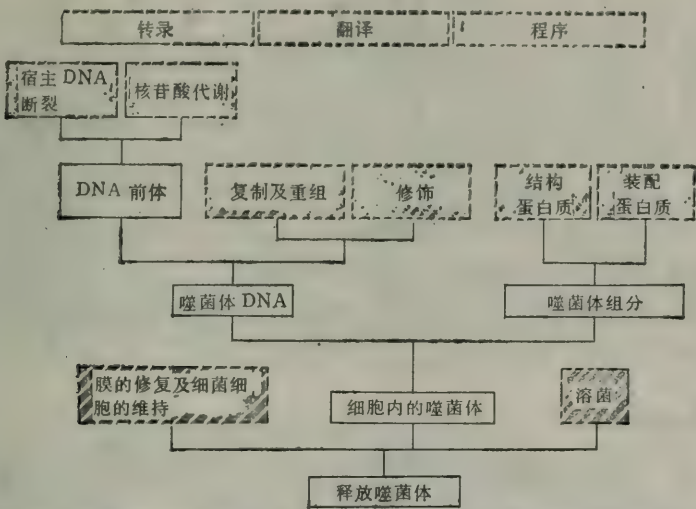


图 15-13 T4 噬菌体基因按功能的分类(虚线框内)。产物及一般功能列在实线框中,连线表明各种各样的功能和成分是怎样地结合起来而产生出许多不同的产物的。

## T4 生活周期概述

- T4 的整个生活周期见图 15-14,现总结如下,温度为 37℃,时间单位为分。
- $t = 0$ , 噬菌体吸附在细菌的细胞壁上,其 DNA 的注入发生在吸附后的几秒内。
  - $t = 1$ , 宿主 DNA、RNA 和蛋白质的合成完全停止。
  - $t = 2$ , 第一个 mRNA 开始合成。
  - $t = 3$ , 细菌 DNA 开始降解。
  - $t = 5$ , 噬菌体 DNA 的合成启动。
  - $t = 9$ , “晚期” mRNA 开始合成。
  - $t = 12$ , 出现完整的头尾。

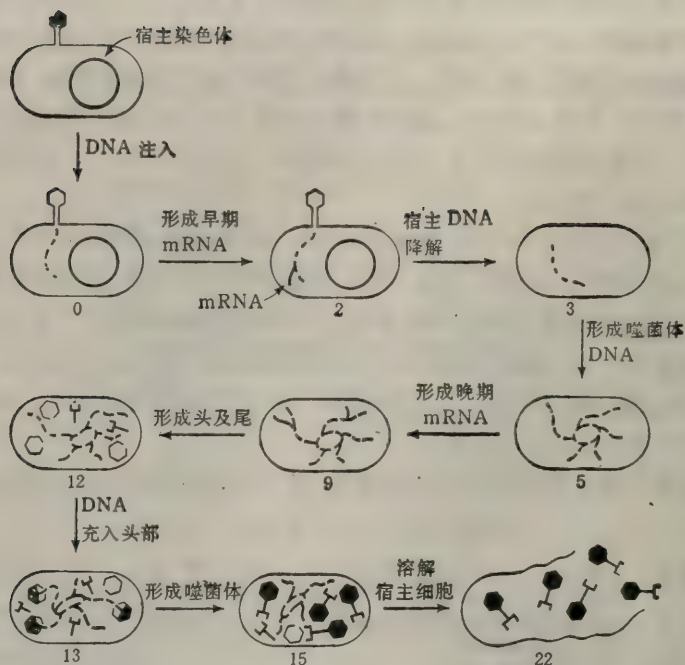


图 15-14 T4 噬菌体的生活周期图解。数字表示注入 DNA 后的时间。为清晰明了起见，mRNA 只在它合成开始时才画出来。

$t = 15$ ，出现第一个完整的噬菌体周期。

$t = 22$ ，细菌裂解，释放出约 300 个子代噬菌体。

这里应注意各个过程的顺序。图 15-15 的电镜照片也部分说明了这个序列。后面将详细介绍其中的几个步骤。

## 细菌大分子合成的关闭

感染后不久，一些能有效地阻止宿主 DNA 转录和复制的噬菌体蛋白质就合成出来了。如宿主 RNA 聚合酶被修饰后几乎不能识别启动子，而且第一个噬菌体 mRNA 编码的 DNase 可迅速降解宿主 DNA，使得细菌 DNA 几乎全部被降解为核苷酸，不能再作为复制和转录的模板了。但是宿主 DNA、RNA 和蛋白质合成的关闭是发生在这些核酸酶的合成之前，有些还是出现在不能合成核酸酶的噬菌体侵染的大肠杆菌，因此，还需要另一种解释才能说明宿主大分子合成的抑制过程。通过遗传学和生物化学方法可以制备一些缺乏 DNA 的噬菌体——幽灵噬菌体 (ghosts)，它们吸附于大肠杆菌后，同样能使宿主大分子的合成关闭，说明这种抑制作用不需要吸附后迅速合成的任何产物。有证据表明，在被完整的噬菌体或幽灵噬菌体吸附后，细胞壁和膜释放一些可溶性物质进入周围的环境。有一种未经证实的假设，即 T4 尾部的吸附改变了大肠杆菌的细胞膜，进而影响了宿主大分子的合成，这可能是由于改变了细胞膜的通透性或影响了宿主染色体与膜的结合，而这两者的结合被认为是宿主复制和转录所必需的。



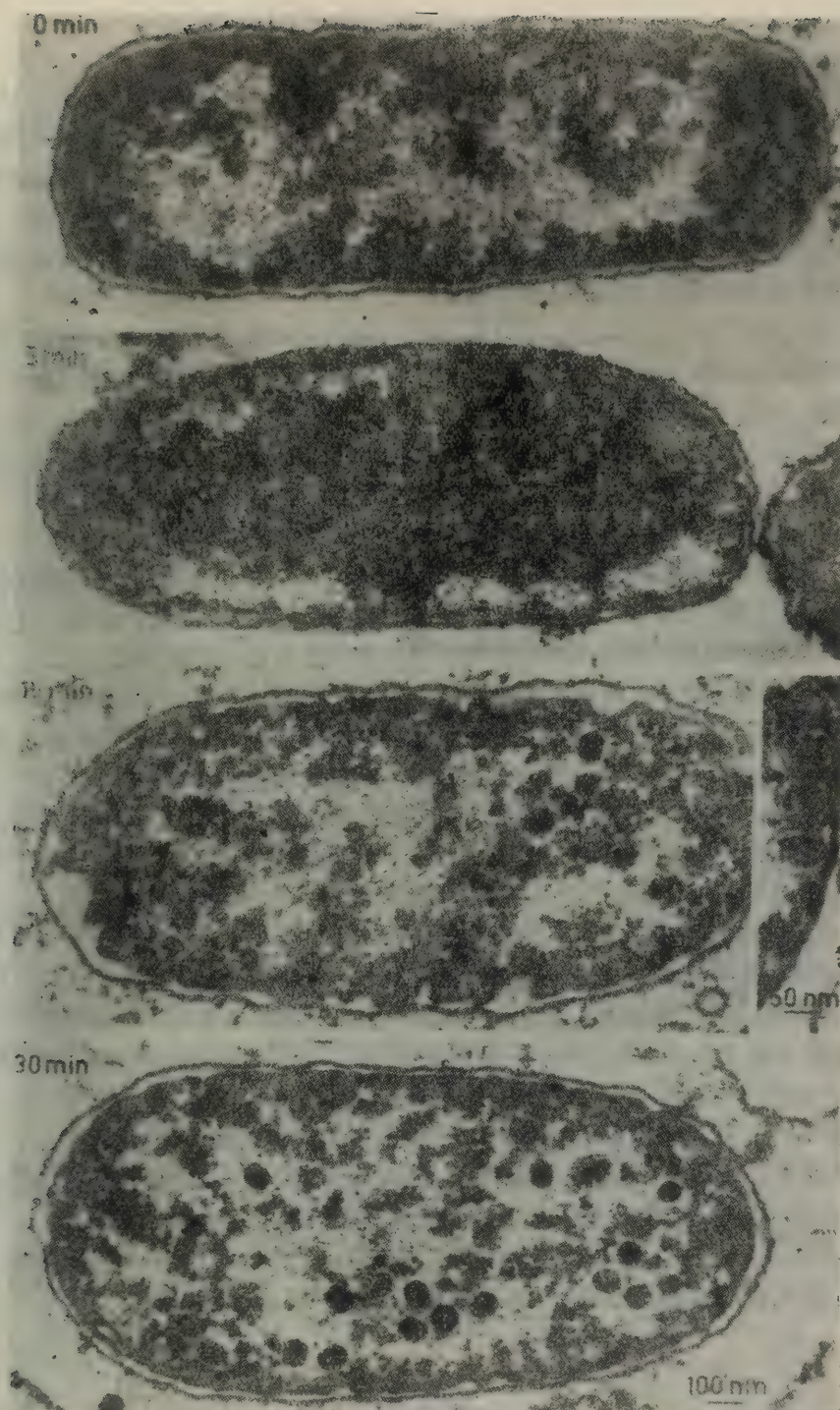


图 15-15 T4 噬菌体发育的电镜照片。照片中所示时间表明于 30℃ 感染后的时间。0 分：感染后立即摄影，表现出未受感染的细胞形态。亮区为含 DNA 的类核部分。5 分：噬菌体诱使核破裂。在这个阶段，发现宿主 DNA 及其断裂产物主要分散在靠近细菌细胞壁处。15 分：约在感染细菌 8 分钟后，噬菌体的用于装配及成熟的结构蛋白质开始形成。再过几分钟，显现出第一个噬菌体颗粒。此后，噬菌体按大约 5 个噬菌体颗粒/分钟的速度逐渐地呈现出来。大约 7 分钟装配出一个颗粒。首先形成一个不含或几乎不含 DNA 的空头部。相似于细菌的核 DNA 的细丝状物质(亮区)是大量的噬菌体 DNA。30 分：出现许多装配完成的头部。因为这是宿主细胞的超薄切片，因此每个细菌细胞中的噬菌体颗粒的实际数目为图中数目的 20 倍 (承 Edward Kellenberger 提供)。

## T4 噬菌体 DNA 的转录

噬菌体蛋白质合成时间的调控是很重要的。例如,若 T4 溶菌酶在生活周期的早期就大量合成,可能在子代噬菌体形成前细菌就裂解了。再如,若在 DNA 复制开始前就装配头部,就有可能将注入的噬菌体 DNA 包装起来而不能增加噬菌体的数量。这种调控的主要机理就是控制 mRNA 的合成,在 T4 噬菌体中,调控过程还包括改变大肠杆菌 RNA 聚合酶使之识别专一的启动子并越过终止子。

T4 DNA 可被转录产生多种 mRNA 分子,各类 mRNA 在不同时间开始转录,在一定的时间间隔内完成。我们将讨论主要的两类:早期 mRNA (DNA 合成前转录) 和晚期 mRNA (复制后转录)。

1. 早期 mRNA 分子编码有关 DNA 合成、mRNA 合成调控、遗传重组中的酶和蛋白质因子及各种非必需蛋白。

2. 晚期 mRNA 分子含有噬菌体结构蛋白、粒子的装配和裂解的信息。

T4 DNA 早期转录的完整过程包括十多种早期 mRNA 的合成,其中多数为重叠的序列,并有几个衰减子(类似于大肠杆菌 *trp* 和 *his* 操纵子中的衰减子)。符合上述基本过程的噬菌体借助于大肠杆菌 RNA 聚合酶(这是细胞中唯一的聚合酶),早期 mRNA 的转录起始于一组启动子;被激活或由于噬菌体专一性蛋白质的加入使此 RNA 聚合酶受到修饰,它不再能识别宿主启动子,而只能识别噬菌体启动子,于是开始晚期 mRNA 的转录。有效的修饰可对多种 T4 早期 mRNA 合成提供暂时性调控。

下面可以见到其他一些种类的噬菌体,最终可使大肠杆菌 RNA 聚合酶失活并产生出噬菌体编码的聚合酶。

T4 编码的蛋白有效地修饰大肠杆菌 RNA 聚合酶活性,其主要步骤如下:

1. 未被修饰的大肠杆菌 RNA 聚合酶的转录作用。感染后不久,第一个 T4 mRNA 分子由宿主 RNA 聚合酶合成。

2. 转换作用。T4 噬菌体头部含几种小分子蛋白,即内部蛋白,在感染后不久,随 DNA 一起注入细菌。其中的一种——转换蛋白,把一种小分子:ADP-核糖(即 ADPR) 连接到聚合酶的一个  $\alpha$  亚基上(图 15-16),从而修饰 RNA 聚合酶。

3. 修饰作用。转换发生后,宿主的 RNA 聚合酶很快地被修饰,也就是加入了一个以上的 ADPR。据认为转换和修饰作用降低了 RNA 聚合酶与  $\sigma$  亚基结合的能力,这样就开始了利用宿主启动子的转录过程。但这些可能都是非必需的功能。

4. 专一性变化。一种噬菌体编码的蛋白质:Mot (可能是 DNA 结合蛋白),降低 RNA 聚合酶识别早期启动子的能力并使之结合到下一组启动子上。

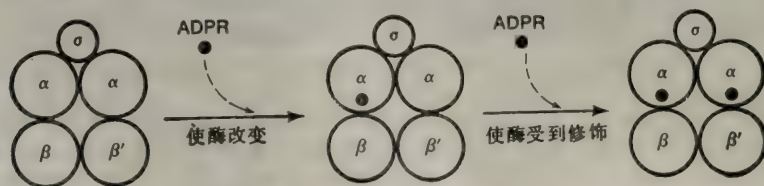


图 15-16 大肠杆菌 RNA 聚合酶由于连续地加入两个分子的 ADPR 而受到改变与修饰。



净结果为按照时间的顺序合成早期 mRNA。RNA 聚合酶专一性发生变化表明, 当不再需要某些基因 mRNA 连续地合成它的表达产物时, 就不再生成这些 mRNA。新的 mRNA 分子将应运而生。

专一性 mRNA 合成终止不久, 由于 mRNA 被快速降解, 被它编码的蛋白质也不再存在。那些长期需要的蛋白质的合成可通过连续性合成 mRNA 或通过抵抗该编码 mRNA 降解的两种途径来实现。前者是一种普遍的途径。

其他的修饰作用能产生更多类型的早期 mRNA。

当晚期 mRNA 开始合成时, RNA 聚合酶成为一类极复杂的酶——其两个  $\alpha$  亚基携带两个 ADPR 分子及四个修饰性蛋白。有时, 称它为“超修饰”(hypermodified) 的 RNA 聚合酶。然而, 这个酶本身是不能使晚期转录得到起始的。晚期 mRNA 仅仅在 DNA 复制开始后才会形成; 在突变体中, DNA 不能复制, 所以, 也从不发生晚期转录作用。因此可以理解, 这种程序化的类型是非常有效的, 因为直到复制作用达到一定阶段之后才需要晚期 mRNA 所编码的一些蛋白质: 头、尾、装配蛋白及溶菌酶。关于晚期 mRNA 合成时为什么需要有 DNA 复制的机理尚不清楚, 但这种情况也已在体外实验中观察到。例如, 在体外实验中, 无论是把纯化了的正常 RNA 聚合酶或是把纯化了的超修饰的 RNA 聚合酶与从纯化了的噬菌体中分离出的 DNA 放在一起, 都不能合成出晚期 mRNA。这些结果以及其他一些结果都指出, DNA 本身在它作为转录的模板之前, 必然地具有某些由正常复制作用所造成的特定构象。注意, 这种调节作用在本书中的前面部分从未提及。1983 年, 发现了一个可以进行晚期转录的体外实验体系。对这个体系的性质进行研究, 肯定地将使我们对这个令人困惑的现象得以了解。

#### T4 DNA 的复制

T4 DNA 复制中的以下五点是很有趣的: (1) 核苷酸来源; (2) 替代 C (胞嘧啶)

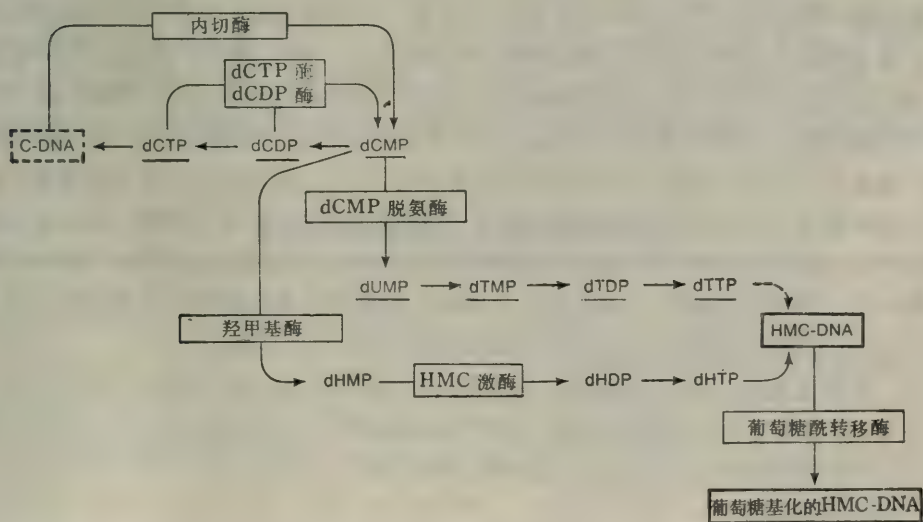
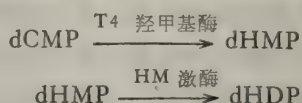


图 15-17 由 T4 噬菌体诱导的 HMC 的生物合成途径。方框中是大分子物质; 细线框表示酶, 粗线框表示 DNA。所发现的只存在于受噬菌体感染的细胞中的物质下面加一横线表示。

的 HMC (5-羟甲基胞嘧啶) 的合成; (3) 阻止 C 的掺入; (4) T4 DNA 的糖基化; (5) 复制作用的酶学。下面将分别讨论。

1. T4 DNA 核苷酸的来源——宿主 DNA 的降解。T4 生活周期的一个早期事件是将宿主的 DNA 降解为脱氧单磷酸核苷酸 (dNMP)。denA 基因和 denB 基因产生的两种内切酶起始上述降解作用。这两种酶只被含有胞嘧啶的 DNA 所活化。它们先切断宿主 DNA 产生双链片段, 这些片段再由基因 46 和基因 47 所控制的外切酶逐步降解为 dNMP。这些单核苷酸再由通常的大肠杆菌酶系催化, 合成出 dATP、dTTP、dGTP 和 dCTP。这些充足的 dNTP 再进一步合成 30 种 T4 DNA 分子, 这些噬菌体 DNA 前体同样也是从头合成 (de novo) 出来的。但为保证提供充足的 dNTP, 合成了在活性上与大肠杆菌酶系相同的五种酶。它们是胸苷酸合成酶、二磷酸核苷还原酶、dCMP 脱氨酶、二氢叶酸还原酶和脱氧核苷酸激酶, 这些酶为非必需的。

2. 5-羟甲基胞嘧啶 (HMC) 的合成。T4 DNA 不含胞嘧啶, 而含 HMC, 但大肠杆菌并不含形成 HMC, 的酶。因此, 由两种噬菌体酶将 dCMP 转变为 dHDP, 此过程如下:



大肠杆菌的核苷酸激酶能形成大肠杆菌中所有的核苷三磷酸, 它再将 dHDP 转变成噬菌体 DNA 中 HMC 的直接前体 dHTP。这些合成途径见图 15-17。

### 3. 阻止胞嘧啶掺入到 T4 DNA 中

从第八章我们已经知道, 由于 dUTP 和 dTTP 分子有相同的 H 键性质, 所以, 大肠杆菌的 DNA 聚合酶不能迅速地区分 dTTP 和 dUTP, 因而还需要其他酶来协助以防止尿嘧啶掺到 DNA 中。同样, 由于 dHTP 和 dCTP 均可与鸟苷形成 H 键, 所以 DNA 聚合酶不能从 dHTP 中识别出 dCTP。T4 内切酶可以作用于含胞嘧啶的宿主 DNA, 因此胞嘧啶决不能存在于 T4 DNA 子链中。并且, 含 C 的 DNA 也不能作为晚期转录的模板, 其原因还不清楚。宿主大肠杆菌并不形成阻止 C 进入 DNA 的酶, 所以一定是噬菌体编码这些酶。若 C 成为子代 DNA 的一部分, dCMP 就可转变为 dCDP 和 dCTP。一种噬菌体的酶, 通常叫做 dCTPase, 即 dCDP-dCTP 焦磷酸化酶, 将 dCDP 和 dCTP 降解为 dCMP。可是形成 dCTP 时要耗用 ATP, 所以这个过程有些浪费。然而, 由于大肠杆菌中只有一种酶 (核苷酸激酶) 负责形成全部的核苷三磷酸, 因而在大肠杆菌中为了阻止形成 dCTP 的反应似乎很难有更经济的过程。而且, 在大肠杆菌中, 绝大多数脱氧核苷二磷酸是由核苷一磷酸通过一种单个的酶所催化的酶促反应形成的 (见第八章)。所以, 在大肠杆菌中阻止 dCMP→dCDP 的反应也是无效的。

另一种噬菌体酶, dCMP 脱氨酶, 可将 dCMP 转变为 dUMP (在合成 dTMP 中), 见图 15-17。这个酶的活性相当于大肠杆菌中类似酶的两倍 (因此是一个非必需基因的产物)。它具有很有趣的节约功能。大肠杆菌 DNA 和 T4 DNA 的碱基组成中含的 (A+T) 分别为 50% 和 66%。在大肠杆菌中, dTTP 和 dCTP 的比例同 DNA 中 T:C 一样, 也是 1:1。当细菌的与噬菌体的 dCMP 脱氨酶共同作用时, dCMP 减少而 dTMP 增加。因此, dTTP:dHTP 为 2:1, 就像 T4 DNA 中的 T 比 HMC 亦为 2:1 一样。



胞嘧啶偶尔也出现在子代噬菌体 DNA 中,假设 T4 内切酶在此种 DNA 合成后不久就将之降解。因为胞嘧啶只存在于每个子代双螺旋的一条链上,所以 DNA 可以通过正常地修复合成途径得到修复。同样,在大肠杆菌 DNA 合成中,尿嘧啶偶尔也插到子代 DNA 中(通过 dUTP 的掺入)。多余的尿嘧啶通常由尿嘧啶- N-糖基化酶 (u-N-glucosylase) 除去(见第八章)。

4. T4 DNA 的糖基化。T4 DNA 中的 HMC 分子给噬菌体带来了问题,因为大肠杆菌中具有一种内切酶,它可以作用于含有 HMC 的特定核苷酸序列,所以噬菌体的糖基化作用就可以避免 T4 DNA 中的 HMC 受这种内切酶破坏。此类糖基化作用涉及到两种噬菌体酶,即  $\alpha$ -糖基转移酶 ( $\alpha$ gt) 和  $\beta$ -糖基转移酶 ( $\beta$ gt)。由于从尿嘧啶核苷二磷酸葡萄糖 (UDPG) 上转移一个葡萄糖到噬菌体 DNA 中的 HMC 上,因此糖基化作用是复制后 (postreplicative) 的修饰作用。大肠杆菌的内切酶不能作用于糖基化的 DNA,可见糖基化作用是一个保护性的手段。简单的遗传学实验表明,保护作用是糖基化作用仅有的基本功能。T4  $\alpha$ gt 突变型不能进行糖基化过程,新合成的 DNA 迅速地被大肠杆菌 HMC 内切酶降解。然而,用一种缺乏 HMC 内切酶的大肠杆菌的突变株 ( $rglB^-$ ) 作为宿主菌时, T4  $\alpha$ gt $^-$  和 T4  $\beta$ gt $^-$  甚至在 DNA 不能糖基化时,也都能正常地生长。

5. T4 DNA 复制作用的酶学机理和复制 DNA 的形式。许多噬菌体基因参与复制过程,其中一些可使大肠杆菌基因的功能加倍。但重建一个含 4 种 dTTP 的体外系统,只能得到 7 种 T4 基因产物,即第 32,41,43,44,45,61,62 号基因的产物,它们的性质见表 15-5。在这个系统中有生物学活性的 T4 DNA 是以纯化了的噬菌体 DNA 作为模板来合成的,合成的速度接近于体内复制及许多正常的细胞内结构(见后)合成的速度,可见这个系统可以再生出体内 (in vivo) 复制所使用的体系。因此, T4 DNA 复制不需任何大肠杆菌复制基因。七种蛋白质组成一个大分子集合体,如基因 43 产物 (T4 DNA 聚合酶),基因 32 产物(螺旋脱稳定蛋白)和 T4 DNA 组成一个稳定的复合体;基因 44 和 62 的产物亦同样,基因 45 蛋白同 44/62 一起,形成依赖 DNA 的 ATP 酶,它的功能多半与大肠杆菌中的解链酶 (helicases) 或 Dna B 蛋白类似。基因 41 产物有引物酶 (primase) 活性,用于前导链的合成,它在此过程中也与基因 61 产物共同配合使用。T4 DNA 复制的酶学机理尚不清楚,但可以认为,对这种体外体系进行认真研究很快就能搞清楚。

表 15-5 T4 噬菌体中 DNA 复制必需的基因

T4 基 因	基因产物的功能
32	螺旋脱稳定作用,单链 DNA 结合蛋白
43	DNA 聚合酶
[44,62]*	结合到聚合酶
45	结合到聚合酶
[41,61]*	引发

\*框中每一对基因产物在分离出来时是紧密结合的复合物。

T4 DNA 复制结构相当复杂,一个原因是可使合成子代 DNA 的速度极高(因而在短时间内产生极大量的 DNA)。在一个生长中的复制叉上,T4 DNA 聚合酶每秒可合成 800 个核苷酸(几乎能与体内的大肠杆菌聚合酶 III 相比拟),而体内 T4 DNA 复制的实际速度更高,为  $5 \times 10^4$  核苷酸/秒·每个细胞。这是由于从两个(或更多个)复制起点进行双向复制,形成复制泡(bubbles),并且在一个复制周期未完全结束时,再在新的复制起点处又起始下一次的复制。结构复杂的第二个原因是 T4 DNA 频繁地发生着遗传重组过程,每个分子的双螺旋经常地发生断裂和重接。人们可以从被感染的细胞中分离出高达亲本分子长度 6 倍的线型连环体(cconcateme) DNA 分子;这些连环体并非产生于滚环复制,而是由于剩余的末端经过遗传重组而形成的。另外,似乎有这种情况,在单一感染的和多重感染的细胞中,在复制的子代分子间发生极为频繁的重组作用,结果细胞中所有未包装的噬菌体 DNA 都是一个含有几百个复制叉的、巨大的、相互连系着的单位中的一个部分(见图 15-18)。这种结构的复杂性严重地阻碍着研究的深入进行。

### T4 噬菌体颗粒的产生

完整 T4 噬菌体颗粒的产生过程可分为两步:装配头、尾及其他结构;包装足够的 DNA,以使噬菌体头中具有略多于一套基因的 DNA。

T4 (以及其他许多噬菌体)装配的研究通过两种技术来完成,这两种技术都需用一大组不能产生完全颗粒的噬菌体突变株。第一种技术是将分别被各种特殊的突变株所感染的不同的细菌培养物进行溶菌处理,然后用电镜检测。该过程指出头部及尾部的合成为相互独立过程,而头及尾的装配则为互相依赖的过程。第二种技术是广泛地采用纯净蛋白质进行互补的测定方法。先得到一个只含有头部物质的噬菌体所感染过的细菌培养物,再得到另一个只含有尾部物质的噬菌体感染过的细菌培养物,分别地再对这两种细菌培养



图 15-18 T4 DNA 复制复合物的电镜照片。注意,它相似于图 6-11 所示的大肠杆菌 DNA 的结构(承 Joel Huberman 提供)。



物进行提取,将两种提取物在体外混合后,可以产生完好的噬菌体颗粒(图 15-19)。“无头”提取物可以被分级分离,从而提取出一种可使“无尾”提取物合成尾部的蛋白质成分。此外,用这种方法还可以分离并鉴定出另一种在尾部装配途径中所需要的蛋白质。

研究表明有两类成分,即结构蛋白质(structural proteins)和形态发生酶(morphogenetic enzymes)。某些结构成分可自发地装配,形成噬菌体结构,而另一些则需要酶的帮助。装配噬菌体头部还需要几种宿主编码的因子。T4 (以及  $\lambda$ ) 的装配途径已经基本上清楚,其具体步骤见图 15-20。

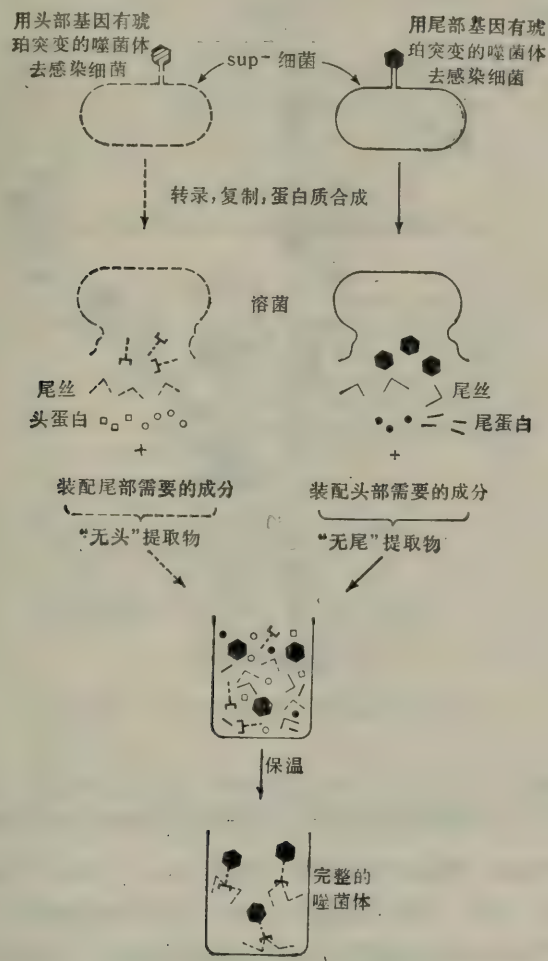


图 15-19 通过体外互补作用产生完整的 T4 噬菌体。

尽管已经可以在体外将 T4 DNA 装配到噬菌体头中,但机理尚不完全清楚。主要的问题就是长链的 DNA 是如何紧密地折叠从而装入噬菌体头部的。据认为,包装过程是 DNA 分子的一端附着到噬菌体头部的一种蛋白质上,接着有一种压缩蛋白或一种小的、基本的分子(如多胺)诱导折叠(见图 15-21)。此过程最明确的一点是 DNA 分子是从长的串联体上一一切割下来的,由于 T4 DNA 的环化位置是可变的,因而,基因切割位点不是在 DNA 中某个特定的、单一的碱基序列上,而只能由可能恰好装入噬菌体

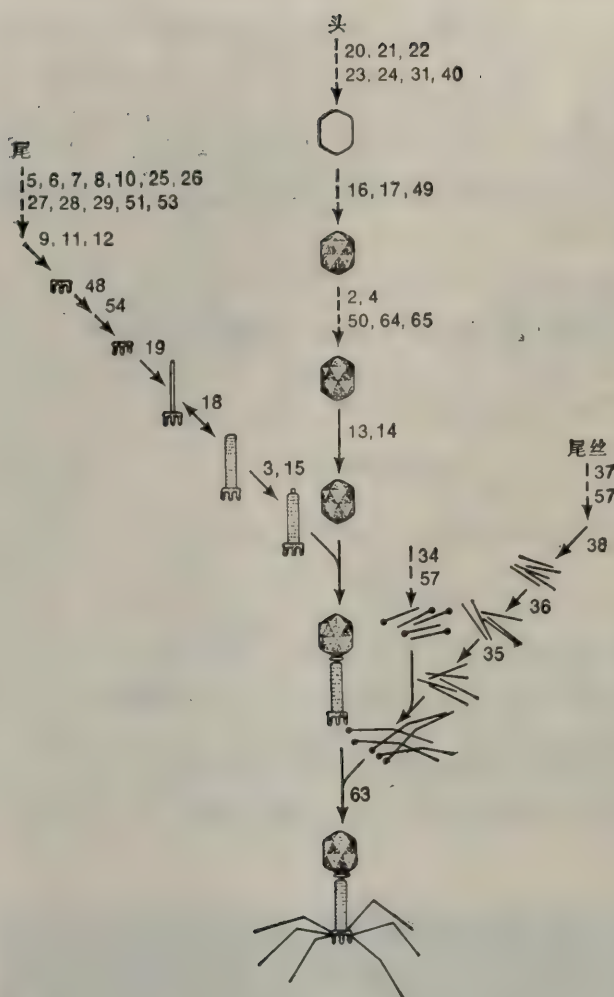


图 15-20 T4 噬菌体的形态遗传学途径(承 William Wood 提供)。

头部的 DNA 的量来决定。“头部充塞机理”假设，先是 DNA 的自由端不断地进入头部，直至头部没有空间，然后 DNA 串联体被切断。这个假设解释了出现终端过剩 (terminal redundancy) 和环状化互换排列 (cyclic permutation) 两种现象的原因 (图 15-22)。最重要的一点是 T4 噬菌体颗粒中所含的 DNA 的量略高于需用来编码 T4 蛋白质的 DNA 的量。所以，当从 DNA 串联分子上切割下一份充塞头部的 DNA 时，最后被包装到头部的 DNA 片段是最早被包装的 DNA 片段的复制品，即被包装的 DNA 的终端是有剩余的。被包装到第二个噬菌体头部的 DNA 分子的第一片段不同于第一个噬菌体头部中第一片段。进一步，因为第二个噬菌体的 DNA 末端也有过剩，所以，第三个噬菌体 DNA 分子当然地从长的串联状的 DNA 分子一个部分起始。因此，由一个受感染的细菌产生的那些噬菌体 DNA 分子是一套环状化互换排列的分子。关于“头部充塞机理”假设，还有两方面重要的证据：(1) 若一个突变噬菌体缺失了一段非必需基因，则



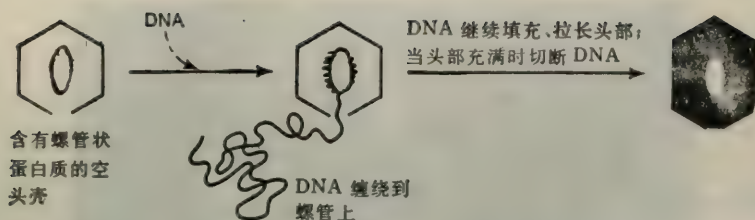


图 15-21 填充 T4 头部的假定模型。切断 DNA 以及重排头部蛋白质发生在这个填充过程的几个不同的阶段。

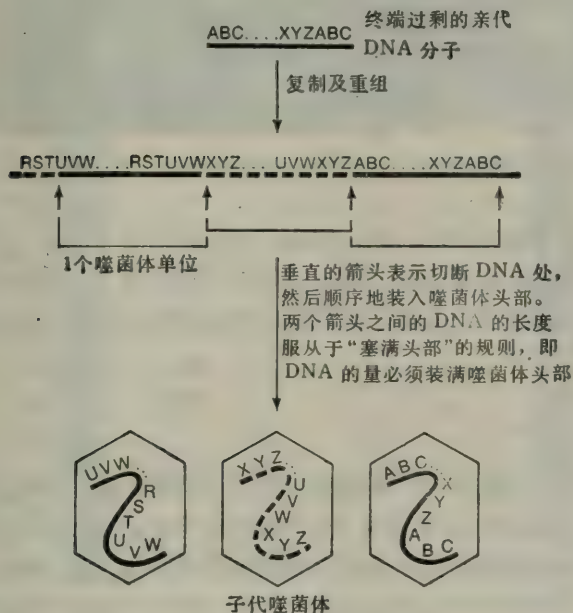


图15-22 T4 DNA 分子环状化互换排列的起点。

终端过剩区域就增加一段和所缺失的 DNA 量相等的片段；(2) 在体积上，0.65,0.8,1.2 和 5 倍于正常噬菌体头部体积的突变噬菌体，其 DNA 的长度也为正常噬菌体 DNA 的 0.65,0.8,1.2 和 5 倍。

## 大肠杆菌 T7 噬菌体

T7 噬菌体是中等大小的噬菌体(表 15-13)，其 DNA 的分子量为  $26 \times 10^6$ ，含有等量的 A,T,G,C 碱基，不含稀有碱基。终端剩余 160 碱基对，但它不是环状化互换排列的序列。DNA 分子被装入头部，一个很短的尾巴附着在头上，见图 15-23。

T7 的基因数目很少，因此对其生活周期进行分析比较简单易行，所以近年来，T7 引起人们的极大兴趣。大部分基因产物经凝胶电泳分离进行了鉴定，其中 34 种已被纯化(1982 年以前)，许多产物已用化学方法测定了序列。最重要的成果是最近已阐明 T7 DNA 的 39 930 碱基对的全部序列，人们可以通过直接观察，对所有的启动子、终止位点、调节

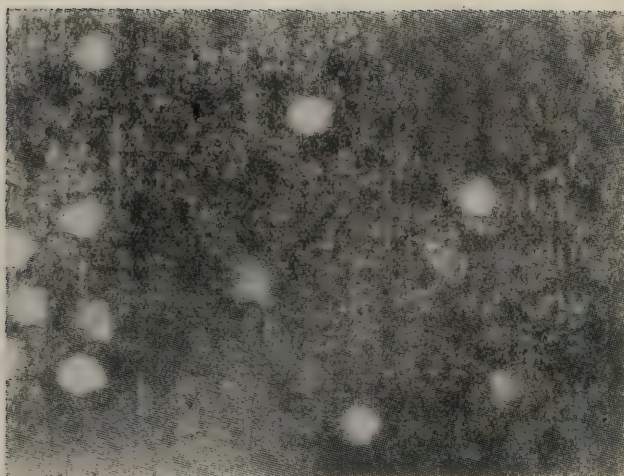


图 15-23 T7 噬菌体的电镜照片。注意,它具有很短的尾部。图中还有三个空着的头部。

位点、间隔区、前导序列、起始密码子、终止密码子和基因等进行定位。实际上,通过在相同的阅读框架中注意 AUG 起始密码子和终止密码子的位置,已经很容易地把以前未知的几个基因检测出来了。进而可以通过每个基因的碱基序列推译出每种蛋白质的氨基酸序列。

### T7 的基因构造

图 15-24 给出了近乎完整的 T7 的遗传学图谱(为简单起见省略了几个未知功能的基因)。在标明的 41 个基因中,已鉴定了 34 个基因的产物,并已知 26 个基因的功能。下面将看到,从基因 0.3 起将按数字的次序顺次地合成出各个基因的产物。

T7 DNA 的量还不到 T4 DNA 的四分之一,可以预计它只复制较少量的宿主基因。对于 T4 来说,至少在 82 个已知的代谢基因中,有 60 个是非必需的。即在所有已知的基因中,非必需的基因占 T4 总基因的 39%,其中大多数和宿主 DNA 的功能相同。然而对于 T7 来说则相反,在已知功能的 26 个基因中,只有 2 个(占 8%)是非必需的。通过测定基因的碱基序列还证明了 T7 DNA 的其他一些较为经济的特点。例如,即使存在着前导序列(leaders)和基因间的间隔区(spacers),这些部分也只有几个核苷酸,而在细菌的遗传系统中这些部分则有几百个核苷酸。T7 有近半数的基因在基因 4 的左边,用于使 mRNA 结合到核糖体上的 Shine-Dalgarno 序列(见第十二章)就在基因 4 左边的一些先导基因(proceeding gene)的编码序列之中。并且一个顺反子(cistron)的终止密码子常常与下一个顺反子的起始密码子重叠,例如序列 UGAUG 终止基因 1.7,并起始基因 1.8,序列 UAAUG 也有双重功能,能对基因 0.5 和基因 0.6 都起作用。还有第三种情况,即相邻基因的编码区是重叠的,其中最大的重叠是 23 个碱基,所以这种重叠比  $\phi$ X174(本章下面介绍)和许多动物病毒(第二十一章)中的重叠小得多。最后,噬菌体生活周期的早期所产生的分解蛋白质比一般细菌蛋白质要小得多,只有 29—883 个氨基酸。

T7 的基因根据其功能而聚集成基因簇,并且按其执行功能的先后排列成一个连续的



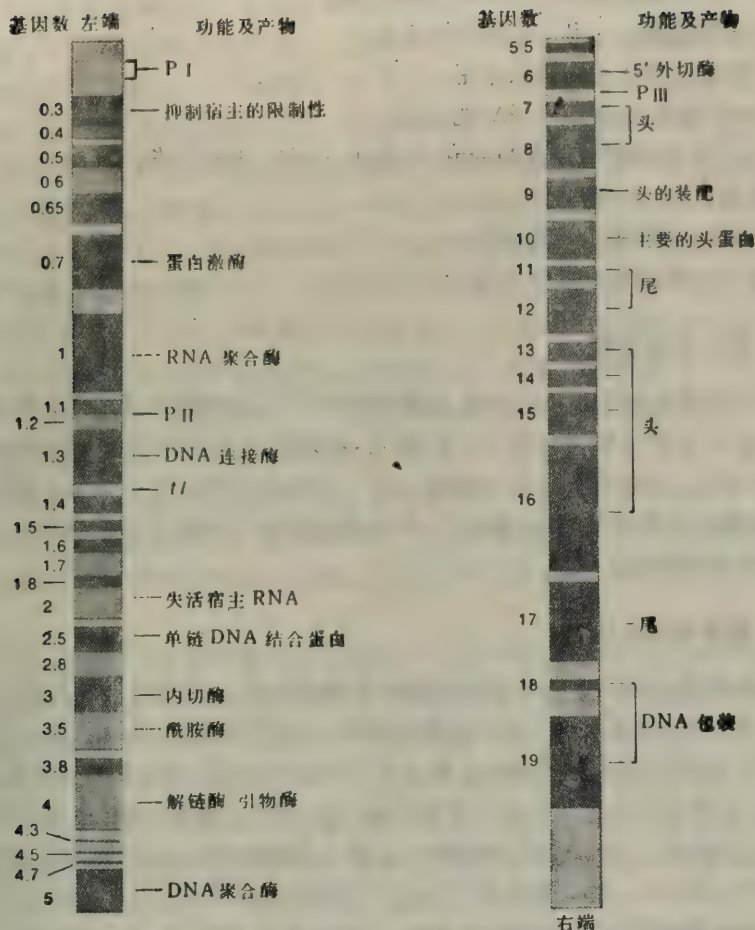


图 15-24 T7 噬菌体的遗传学及物理学图谱。基因均已标以数码。图中没有标明功能的基因，或是并不重要或是功能尚不明了。图中方块的大小是其蛋白质产物大致的比例。方块中阴影的深浅仅仅是用来区分彼此相邻的基因，并无其他含意。启动子用 pI, pII, pIII 表示，终止子用 tI 表示。

次序。第一个基因(在图 15-24 中，在遗传学图谱中从上到下阅读，或按图谱的标准方向从左到右阅读)是感染宿主细胞后噬菌体 DNA 存活所必需的基因；T7 的一个普通的宿主——大肠杆菌 B 株，含有一种限制性内切酶，它可以切断 T7 DNA 中的专一性碱基序列，噬菌体 DNA 中的第一个基因的产物直接地结合在这个酶上，从而抑制这个核酸酶的活性。紧接着是几个控制转录作用的基因，然后是 DNA 复制基因，再后是编码噬菌体颗粒的结构蛋白质的基因，最后则是几个能将新合成的 DNA 装入到噬菌体头部的基因。

## T7 的生活周期

T7 的生活周期总结如下(30°, 时间单位为分钟):

t = 0 噬菌体的吸附。

- $t = 0 - 1$  起始噬菌体 DNA 的缓慢注射。
  - $t = 2$  起始噬菌体 mRNA 合成的起始。
  - $t = 4$  开始关闭宿主的转录作用。
  - $t = 8 - 9$  起始噬菌体 DNA 的合成。
  - $t = 8 - 10$  起始结构蛋白的合成,噬菌体 DNA 注射完毕。
  - $t = 15$  出现第一个噬菌体。
  - $t = 25$  裂解并释放子代噬菌体。
- 下面将对大部分阶段进行介绍。

## T7 DNA 注入宿主细胞

T4 及其他多数噬菌体的 DNA 在吸附后不到 1 分钟就迅速地注入到宿主细胞中,而 T7 约需 10 分钟才能将全部 DNA 注入,其原因还不清楚(其他噬菌体注射机理同样也不完全清楚)。缓慢地注入这一特点在 T7 的生活史中非常有意义,因为一个基因在它被注入之前是不能被转录的。事实上,转录的时间在部分程度上取决于注入过程的动力学,我们将在下面讨论。

## T7 DNA 转录作用的调控

T7 的生活周期通常可分早期和晚期两个转录阶段。早期转录是利用大肠杆菌 RNA 聚合酶,对基因 0.3 至基因 1.3 这一段进行转录。晚期转录是使用一个新合成的 T7 RNA 聚合酶。T7 的转录被短暂地调控。转录子(transcripts)可以分为主要的三类,分别标以 I,II,III(见图 15-25),其中每一类都从同一个 DNA 链即  $r$  链(当 DNA 分子按标准方向画出时,转录方向为向右)开始转录。第 I 类转录子由大肠杆菌的 RNA 聚合酶合成,组成早期 mRNA。第 I 类转录子有三个不同的启动子,它们的位置很近,因而每一个 mRNA 都包含着相同的基因的转录产物。若把这些 mRNA 分子中的每一个都看作是从一个单个的启动子 pI 开始的、被称为转录子 I 的一个 mRNA 分子,那么这对于我们讨论转录的调控将是很方便的。三个启动子的存在很可能仅仅是使 mRNA 的合成的净速度提高 3 倍。

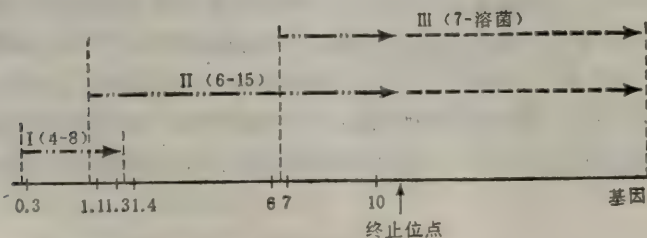


图 15-25 简化了的 T7 转录图谱。DNA 及基因数目用实线表示,三类转录作用用---表示(其上的数字表示时间间隔(37°C,分钟)),在这段时间间隔中完成转录作用。终止位点用垂直箭头表示,大约可有 90% 的终止效率;因此大约有 10% 的第 II 类和第 III 类分子进一步转录,图中用虚线(---)表示。



转录子 I 的合成是在感染后 2 分钟开始的。这段时间的延迟可能是由于 DNA 的缓慢注射。细菌 RNA 聚合酶参与转录,在基因 1.3 之后的地方终止转录。大肠杆菌 RNase III 将转录子 I 切割为 5 个 mRNA 分子,但在实验室条件下此过程不是必需的,因为,如果是缺乏 RNase III 的大肠杆菌突变株,转录子仍保持完整,产生出的噬菌体似乎是正常的。然而在自然条件下或在其他细菌宿主内可能需要有上述加工过程。

转录子 I 可翻译产生几种蛋白质,其中一种是蛋白激酶 (protein kinase, 基因 0.7 的产物),它可使大肠杆菌 RNA 聚合酶磷酸化,而使之失活,从而减少大肠杆菌 DNA 的转录,这是噬菌体取代细菌的第一步。T7 噬菌体对大多数大肠杆菌的实验菌株的感染过程,并不需要这个激酶。然而, T7 噬菌体的基因 0.7 突变株与正常的野生型噬菌体相比,产生一个较小的爆破值,这多半是因为细胞继续地合成着大肠杆菌的 RNA,而大肠杆菌 RNA 对于 T7 无用,因而使形成 T7 颗粒的来源减少。我们还将看到,基因 0.7 后来形成一个大肠杆菌 RNA 聚合酶的抑制物。转录子 I 的第二个翻译产物是 T7 RNA 聚合酶(基因 1 的产物),这是一个必需的酶,它关系到进一步的转录。与大肠杆菌 RNA 聚合酶相反, T7 RNA 聚合酶是一个单链的多肽,而前者是一个多亚基的复合物。比起大肠杆菌来,推测它的转录作用的起始与终止的要求很简单,而大肠杆菌则有较为复杂的要求。

转录子 II 及转录子 III 代表着各种类型的、起始于不同启动子的 mRNA。这些启动子不能被大肠杆菌的 RNA 聚合酶识别,它们只被转录子 I 翻译产生的 T7 RNA 聚合酶识别。但是,这些启动子并没有立刻在 T7 RNA 聚合酶形成时就开始转录,而是有一段滞后的时间。第 II 类转录子(转录子 II)先合成出来,然而它的启动子却比第 III 类转录子(转录子 III)的启动子弱得多。对于第 II、III 类转录子出现的延迟以及这两类转录子的转录在起始时间上的差异,有着同样的解释。这是因为 T7 DNA 缓慢地注射以及因为第 II 类的启动子进入细胞比第 III 类进入细胞早几分钟。

转录子 II 的转录作用起始于第 I 类转录子的终止位点的上游,因此,在转录子 I 及转录子 II 中都含有基因 1.3,原因尚不明,特别是缺少 T7 DNA 连接酶的基因 1.3 突变体,可以在提供大肠杆菌 DNA 连接酶的情况下很好地生长——至少在实验室条件下是如此。在自然界,可能 T7 经常遇到宿主菌株,而宿主菌株不能为迅速生长的噬菌体提供足够量的连接酶,于是演化出能生长于这种宿主中的 T7,或者说 T7 的酶系具有足够的优势,即能使 T7 的拖延时间的 RNA 合成仍然得到保证。T7 RNA 聚合酶可以忽略引起转录子 I 终止的位点,使转录子向右地转录到第二个终止位点(见图 15-25)。终止位点 90% 有效,这意味着有 10% 的转录子不终止,继续前进,直到 DNA 分子的末端处。因为终止位点刚好在基因 10 之后,(基因 10 编码主要的头部蛋白),所以人们认为终止位点的位置有利于增加基因 10 的产物浓度。就在此终止位点之后,又有编码各种尾部蛋白及一些次要的头部蛋白的各个基因,噬菌体对于这些基因的蛋白产物的需求量较低。

第 II 类转录子的基因(基因 II)形成对大肠杆菌 RNA 聚合酶的第二种抑制物。这种抑制物与前面曾提过的基因 0.7 的蛋白产物:蛋白激酶(降低大肠杆菌 RNA 聚合酶活性)结合,就使大肠杆菌 RNA 聚合酶的活性被完全抑制。因此,于感染后 8 分钟不再有大肠杆菌的 RNA 的合成,噬菌体已完成取代细菌的全部过程。这时由大肠杆菌 RNA

聚合酶形成的转录子 I 也不再合成了。停止合成转录子 I 对噬菌体 T7 来说,是十分经济的,因为此时已有足够的 T7 RNA 聚合酶去完成其生活周期,而不再需要这个转录子的产物了。

转录子 II 的 DNA 还含有参与 DNA 复制的各种酶和蛋白质因子的基因以及形成结构蛋白的其他一些基因。事实上,在对复制作用的基因转录以后不久,就开始转录结构蛋白的基因。初看似乎转录的量不够,因为噬菌体颗粒的成熟前的装配过程及噬菌体 DNA 的包装过程可能会导致复制过程终止或至少大大地减少复制作用的模板数。这种安排与 T4 中所见到的过程相反, T4 结构蛋白的合成因为涉及复制的酶系是延滞的。然而在 T7 中,虽然它的 DNA 分子很少,但它可以迅速地复制(约 20 秒一次),一些与成熟作用有关的蛋白质(基因 18 及基因 19 的蛋白产物)则翻译得晚,头部的装配过程很慢;在复制作用合成出许多新的 DNA 分子之前,实际上没有 DNA 分子的包装过程。

一旦合成出 DNA 复制作用所需要的酶系,就应当不再需要对这个区域进行转录,因为对于大部分情况来说,已经具有复制作用所需的各种蛋白质。然而,使转录子 II 的合成结束,也就停止了结构蛋白基因的转录,但是噬菌体对这些结构蛋白基因的产物的需要量是符合化学计算的。所以,通过翻译两种晚期 mRNA 分子(II 及 III)以及停止合成转录子 II 来达到对效率及对连续合成结构蛋白的要求。人们认为,如果转录子 II 在遇到转录子 III 的启动子之前就终止转录,也会得到与上面相同的结果,但是自然界的进化选用了另一条途径。

在感染后 15 分钟转录子 I 的合成停止。这种关闭作用带有普遍性意义,它关系到转录子 II、III 的启动子,影响启动子 pII 和 pIII 的相对强度。于是,转录作用全面地减弱,仅仅较强的启动子 pIII 还保留活性。转录活性全面减弱的机理尚不清楚,但是已知这个过程需要基因 3.5 的蛋白产物(N-乙酰胞壁酰-L-丙氨酰-脱酰胺酶)。这个蛋白质是一种类溶菌酶的酶,对细胞溶解也有作用。

## T7 DNA 复制

T7 DNA 复制期间所专一性使用的一些蛋白质是转录子 II 的翻译产物。基因 3 及基因 6 编码两种核酸酶,它们的作用是打断宿主的 DNA。此作用为噬菌体 DNA 合成提供大部分核苷酸。基因 5 的蛋白产物为 T7 DNA 聚合酶,它缺乏 5'→3' 外切酶的活性,这种活性改由基因 6 蛋白提供。基因 4 的蛋白产物有引物酶及解链酶两种活性。基因 2.5 编码单链 DNA 结合蛋白。

T7 DNA 复制的模式详述于第八章(见图 8-46 至图 8-48)。推荐读者读一读该部分,最重要之点为原初的亲代分子产生一个长的,线性串联体。终端剩余单位的分子从这个串联体上切下,切断的方式将在后面介绍。

## T7 噬菌体的成熟

T7 DNA 包装到噬菌体中的过程与 T4 DNA 的包装过程很不相同,例如,(1) T7 DNA 也有末端冗余现象,但是没有环状化互换排列现象;(2) 凡具有遗传缺失的突变株噬菌体所含的 DNA 少于野生型噬菌体的。T7 成熟的机理尚不清楚,但有人提出一个



方案（见图 15-26），其要点是通过在 3'-OH 及 5'-P 终端产生一个缺刻（nick），串联体上的 DNA 对一些酶的作用变得敏感。这些酶可以形成在 T7 噬菌体颗粒中所见到的有平整末端（blunt-ended）的末端冗余的 DNA 分子。

关于 T7 头部、T7 尾部的形成以及 T7 体外装配的详细过程已经了解清楚，可以阅读本章末列出的参考文献。

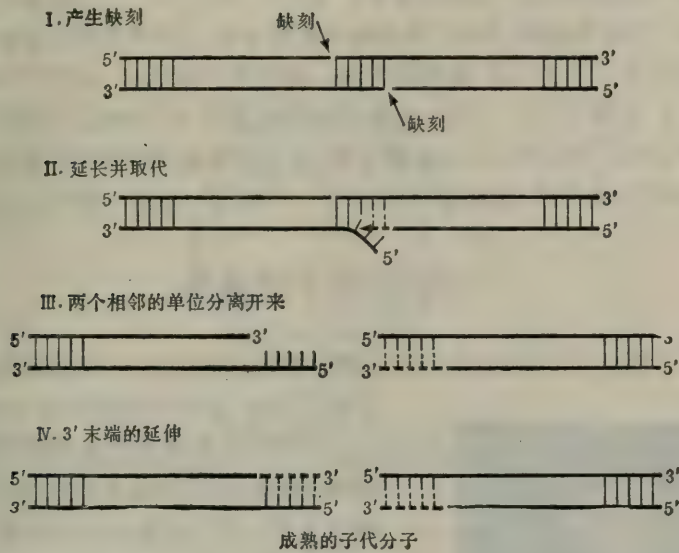


图 15-26 从串联体上产生出成熟的 T7 DNA 的假设图解。三处碱基配对区表示三个冗余的 DNA 区域。步骤如下：I. 在每一个冗余区的末端形成缺刻，产生出两个 3'-OH 末端。II. DNA 聚合酶将核苷酸加到由缺刻作用而生成的 3' 末端上，取代原先的 5' 末端亲代链。III. 当上述取代反应进行到第二个缺刻时，两个双链片段就分离开来，产生一个完全的子代分子。IV. 另一个尚不完全的子代分子的 3' 末端由 DNA 聚合酶进行延展，最后形成第二个完全的子代分子（由 J. D. Watson 提出的模型）。

## 两种枯草杆菌噬菌体

这两种噬菌体，PBS-2 和 SPO1，具有和噬菌体 T4、T7 相关的有趣的调节性能，下面我们将简述之。

### 噬菌体 PBS-2：不需要宿主的聚合酶

到目前为止，我们讨论的噬菌体由于在感染时没有其他 RNA 聚合酶，所以都要利用宿主的 RNA 聚合酶来起始转录作用。对于一个噬菌体来说必须建立起一套转录体系使本身的 DNA 比宿主 DNA 得到优先转录。一般，是合成一个新的 RNA 聚合酶（T7）和/或修饰宿主的 RNA 聚合酶（T4 和 T7）。但是枯草芽孢杆菌（*Bacillus subtilis*）噬菌体 PBS-2 则有不同的机理。它合成一个专一于噬菌体的 RNA 聚合酶，该酶被包装在噬菌体头部并与噬菌体 DNA 一起注射到宿主细胞中，这种现象在噬菌体中很少见到，但正如将在第二十一章中所述，这种情况在动物病毒中并不罕见。

## 噬菌体 SPO 1

枯草芽孢杆菌噬菌体 SPO 1 的调节过程兼有 T4 和 T7 的特点。它有起始于三组不同启动子 pI、pII、pIII 的三类 SPO 1 转录子。枯草芽孢杆菌的 RNA 聚合酶是一种与大肠杆菌的聚合酶类似的多亚基蛋白质,它不从 pII 或 pIII 而只从 pI 开始转录。基因 28 的蛋白产物是由转录子 I 翻译而得。它是一种  $\sigma$  因子,可以取代宿主酶上的  $\sigma$  亚基。被基因 28 产物修饰过的 RNA 聚合酶不能再在宿主的启动子和枯草芽孢杆菌噬菌体的 pI 处起作用,但它可在 pII 处起始转录。另外两种蛋白质,基因 33 和基因 34 的蛋白产物是从转录子 II 转译的。它们在一起可以取代基因 28 的蛋白产物的作用。而形成第二个修饰过的 RNA 聚合酶,仅仅能在 pIII 处起始转录。通过这些 RNA 聚合酶修饰物(即  $\sigma$  亚基)的顺次级联合成,使噬菌体得到了一个早、中、晚期 mRNA 的有序合成程序。

## 大肠杆菌 $\lambda$ 噬菌体

大肠杆菌  $\lambda$  噬菌体(见图 15-27)有两种交替的生活周期,即裂解与溶源。本章我们仅研究裂解,溶源周期将在第十六章讨论。



图 15-27  $\lambda$  噬菌体的电镜照片(承 Robley Williams 提供)。

目前,  $\lambda$  噬菌体可能是研究得最透彻的双链 DNA 噬菌体,尤其是关于它的转录的调节过程的精巧方式。对  $\lambda$  噬菌体的详细研究已对分子生物学的发展做出了许多重大贡献。此处仅举几个例子,如双向复制、转录终止、Rho 蛋白、抗终止蛋白、DNA 连接酶、DNA 解旋酶 (gyrase)、位点专一的重组 (site-specific recombination)、环状 DNA 分子、断裂和再接合遗传重组的 (break and rejoin genetic recombination)、SOS 修复等的发现。进一步,  $\lambda$  噬菌体还作为人们用于了解许多其他生物学过程的体系。

### $\lambda$ 噬菌体 DNA 及其注入后的转变

$\lambda$  噬菌体是一种中等大小的噬菌体(表 15-3), 分子量为  $31 \times 10^6$  (50 000 碱基对), 是双链 DNA 分子。该 DNA 分子具有特殊结构——在双链分子的两端, 5' 末端比 3' 末端多 12 个碱基。两个单链末端区域粘性末端的碱基顺序是彼此互补的(见图 15-28)。因此, 通过这两个单链末端形成碱基对, 线性 DNA 分子可以环化, 产生含有两个单链缺口的不闭合环。在此环化过程中, 新形成的双链区中不缺失任何碱基, DNA 连接酶可将其转变为共价的闭合环。这个环化过程在实验室条件下很易进行, 在被感染的细菌中亦可按下述时间顺序进行(37℃, 分钟):

$t = 0$ , 噬菌体吸附, 线性 DNA 分子注入。

$t = 1-2$ , 不闭合环形成。

$t = 3-6$ , 大肠杆菌 DNA 连接酶使环闭合, 大肠杆菌 DNA 解旋酶使之形成超螺旋。



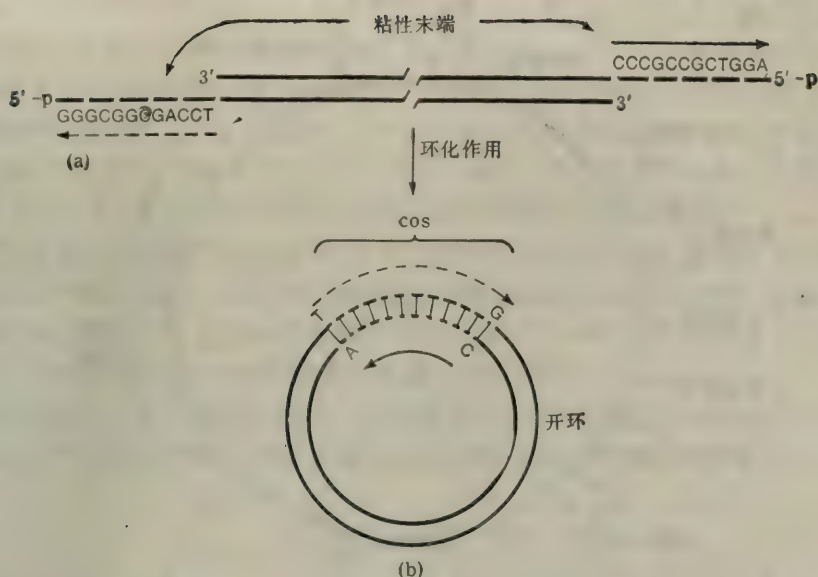


图 15-28  $\lambda$  噬菌体 DNA。(a)  $\lambda$  噬菌体 DNA 分子的线性图式, 示出互补的单链末端(粘性末端)。注意, 12 个碱基中有 10 个是 G 或 C。(b) 通过粘性末端中的碱基进行配对, 使线性分子发生环化作用。新形成的双链区叫做 cos 区。

旋。

## $\lambda$ 基因的构造和调节位点

现在已鉴定出 46 个  $\lambda$  基因, 其中有 14 个对裂解过程是非必需的, 只有 7 个基因对裂解和溶源两者都是非必需的。通过凝胶电泳, 已分离或纯化了大部分  $\lambda$  蛋白质。 $\lambda$  DNA 的许多区域及大部分调节位点、启动子以及终止位点的碱基序列已经测定出来, 从蛋白质和 DNA 分子的大小来估计可以知道, 被人们鉴定过的基因已占全部基因的 90%。

$\lambda$  噬菌体的遗传图谱见图 15-29。待讨论过溶源过程之后, 就可以理解该图谱。它通常被画为环形。和噬菌体 T7 一样,  $\lambda$  噬菌体的基因也按功能聚集成簇, 如头、尾、复制和重组基因分别地形成四个不同的基因簇。许多  $\lambda$  噬菌体蛋白质, 例如调节蛋白以及与 DNA 合成有关蛋白质, 在 DNA 上的特定部位起作用。一般说来, 编码这些蛋白质的区域都相邻于它们的作用位点(当只有一个作用位点时), 例如, DNA 复制起始点 (origin) 处在基因 O 的编码序列中, 而基因 O 就是 DNA 复制起始蛋白的基因。

## 裂解周期的时间顺序

$\lambda$  裂解周期的时间过程极其复杂, 这很可能是由于某些基因既用于裂解又用于溶源所致大致顺序如下(37°C, 分钟):

- $t = 0$  噬菌体吸附和 DNA 注入。
- $t = 3$  前早期 mRNA 合成(即 pre-early, 第一个 mRNA 的合成)。
- $t = 25$  早期 mRNA 合成(两类早期 mRNA 合成)。
- $t = 6$  DNA 复制开始。

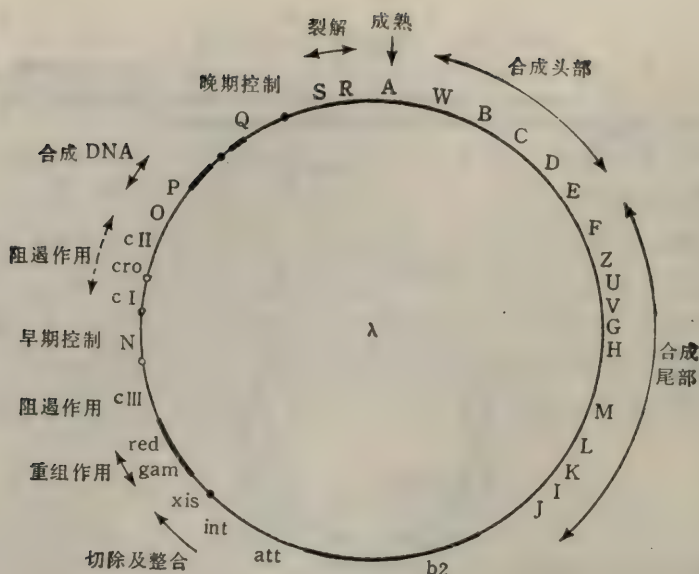


图 15-29  $\lambda$  噬菌体的遗传学图谱。图中调节基因及调节功能的下面用一小横线表示。图中并未显示所有的基因。主要的调节位点用黑色圆点表示。对于溶菌和溶源两个循环均不重要的 DNA 区域用粗黑线表示。

- $t = 9$  晚期 mRNA 合成开始。
- $t = 10$  结构蛋白开始生成。
- $t = 22$  第一个噬菌体颗粒完成。
- $t = 45$  裂解和释放子代噬菌体。

请注意,循环周期为 45 分钟,而 T4 和 T7 是 22—25 分钟。这段时间较长,反映出具有两种交替的生活周期的噬菌体需要有较为复杂的调控途径。

### 维持宿主体系完整性的必要条件

T4 和 T7 生活周期的一个基本特征就是迅速杀死宿主并溶解宿主 DNA。而  $\lambda$  噬菌体则不同,在其溶源周期中,其宿主必须存活下去。确实有  $\lambda$  的基因产物缓慢地减弱宿主大分子的合成进而杀死宿主细胞,但这些产物是非必需的,即,若这些  $\lambda$  的基因在遗传学上缺失, $\lambda$  仍然能正常地发育,其爆破值 (burst size, 见图 15-6) 也并不明显地减少(这些基因产物是如何工作的,在  $\lambda$  的生活周期中有何作用尚不清楚,本书不做讨论)。最后,裂解相 (lytic phase) 中受感染的细胞的 DNA 复制系统被破坏的结果就是细胞死亡并裂解。

本章后面我们还将讨论一种噬菌体 (M13), 它由被感染的细胞连续地生成,但它永不杀死该细菌;所以,宿主细胞甚至可以成为产生噬菌体的机器,而不丧失其本身存活与复制能力。

### $\lambda$ 的转录顺序

在前述的噬菌体中,各种 mRNA 分子的合成时间主要由决定可以供给多少个启动



子的机理所限定,即由新的 RNA 聚合酶的合成、宿主 RNA 聚合酶的修饰或启动子进入宿主细胞内的时间等决定。在 $\lambda$ 中,宿主的 RNA 聚合酶亦被修饰,但不是为了去识别噬菌体的启动子,而是修饰作用使得聚合酶能通过终止位点。事实上,晚期 mRNA 合成受阻(延滞)的主要原因是转录和翻译编码抗终止蛋白的基因需要一定时间。图 15-30 是 $\lambda$ 的遗传学图谱,它包括三个调节基因: *cro*、*N* 和 *Q*; 三个启动子: *pL*、*pR* 和 *pR2*; 五个终止位点: *tL1*、*tR1*、*tR2*、*tR3* 和 *tR4*; 七个 mRNA 分子。其中 L 和 R 系列分别从互补的双链 DNA 分子左向和右向转录,希望读者在下面的讨论中参考图 15-30。

既然不需要宿主丧失功能(见前面),可预见转录的方式如下。首先是 *O*、*P* 基因被转录,它们的产物为 DNA 合成所必需,接着是转录编码结构蛋白的基因,最后是产生包装体系和裂解蛋白。以上这些是最基本的,同时形成另外一些小转录子,它们分别编码一些调节蛋白,使转录在适当时候开启或停止,起到修饰作用。

转录序列见下文,下面基因、位点、mRNA 分子的意义见图 15-30 和表 15-6。在相同的或接近的时间的开始的两个过程,使用相同的数字,如 1A 和 1B。

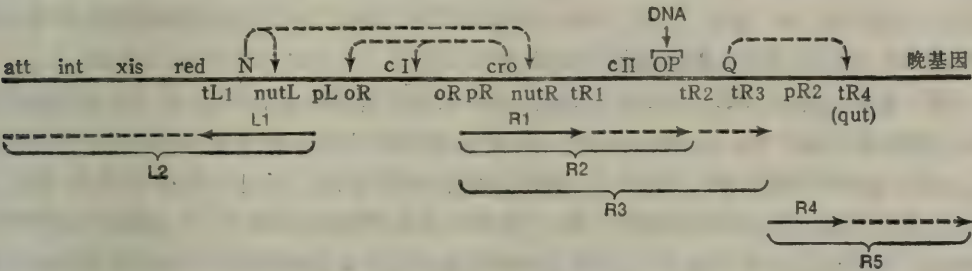


图 15-30  $\lambda$ 噬菌体的调节基因的遗传学图谱。基因都写在黑线上面,位点写在黑线下面,上面的虚线箭头表示N蛋白、Cro 蛋白、Q 蛋白的作用位点。mRNA 分子用虚线表示。

表 15-6  $\lambda$ 噬菌体中的某些位点及某些基因

位点或基因产物	描 述
位点:	
oL,oR	左向或右向操纵子
pL, pR	左向或右向启动子
tL(1, 2)	左向转录作用的终止位点
tR(1,2,3,4, 5)	右向转录作用的终止位点
基因产物:	
Cro 蛋白	对于由 pL 和 pR 开始的转录过程进行抑制的蛋白质
cII 蛋白	推迟晚转录的蛋白质(在溶源途径中起其他的作用,如十六章所述)
N 蛋白	抗终止作用的蛋白质,作用于 tL1, tR1 及 tR2
O,P蛋白	DNA 复制所需要的蛋白质
Q 蛋白	抗终止作用的蛋白质,作用于 tR4
R	右向合成出的 mRNA
L	左向合成出的 mRNA
R4	构造型合成的 mRNA

1A. 转录子 L1 的合成。起始于启动子 *pL*, 终止于位点 *tL1*。此转录子只编码基 因N的蛋白,N蛋白是一个允许 DNA 的某些区域转录的正调节因子。

1B. 转录子 R1 的合成。起始于启动子 pR, 终止于位点 tR1。基因 Cro 编码 Cro 蛋白, Cro 蛋白是一个负调节因子, 它最终关闭从 pL 和从 pR 起始的转录。

2. 转录子 R2 的合成。tR1 是一个弱的终止位点, 约有一半的转录子 R1 mRNA 一直延伸到终止位点 tR2 处才停止转录。Cro 蛋白(功能见 4) 从这两类 mRNA 分子上转译出来。不在 R1 内的 R2 部分包含着基因 O 及基因 P, 它们的产物在 DNA 复制中是必需的, O、P 的产物在此时由 R2 转译出来, 但它们的浓度只够 DNA 复制一次所用。

3A. 在 tL1 位点上的抗终止作用和转录子 L2 的合成。基因 N 蛋白一旦被合成, 就与一个宿主蛋白质(称作 NasA 蛋白) 结合到一起, 去修饰大肠杆菌 RNA 聚合酶, 使之不在 tL1 处终止。这样, 转录子 L1 不再生成, 但可延伸而形成转录子 L2。L2 mRNA 转译, 产生出几种蛋白产物, 它们都是非必需的, 但使得形成噬菌体的效率更高, 这些蛋白质的功能尚不清楚。对 N 蛋白所参与的抗终止作用的机理的探讨仍为一个活跃的研究领域。这时显然还需要  $\lambda$  DNA 上的一个被称为 nut 的位点 (N-utilization), 与 N 蛋白的利用有关, 此位点处于 pL 的“下游”(down stream)。尽管目前的证据很少, 但是 N 蛋白有可能结合在 nut 位点上, 在 RNA 聚合酶沿着 DNA 分子运动, 形成转录子 L1 和转录子 R1 时, 被 RNA 聚合酶识别出来。

3B. 在终止位点 tR1 处和终止位点 tR2 处的抗终止作用、转录子 R3 的合成及 DNA 复制蛋白的产生。N 蛋白一旦合成, 被修饰过的 RNA 聚合酶就不在 tR1 处终止, 也不再形成转录子 R1, RNA 聚合酶也越过 tR2 位点, 而延长转录子 R2, 使它在终止位点 tR3 处停止, 形成转录子 R3。即使 RNA 聚合酶被修饰, 它也得在 tR3 处停止转录。tR3 位点对于 N 蛋白的抗终止效应不敏感, R3 的合成提供足够的 mRNA, 使 O、P 蛋白的浓度足以用于 DNA 的复制。cII 蛋白也是由转录子 R2 和转录子 R3 翻译出来的蛋白质, 其功能是延迟晚期 mRNA 的合成。

4. 关闭转录子 L2 的合成。当前面的现象发生时, Cro 蛋白的浓度已经开始增加。Cro 蛋白为一种阻遏蛋白, 它结合到左向操纵子 oL 上, 阻断从左向启动子 pL 处开始的转录。因此, 在这一时刻, 已形成足够量的 N 蛋白, 不再需要转录子 L1 及 L2, 也就不再去合成 L1 及 L2 了。

5A. 开闭转录子 R3 的合成。Cro 蛋白阻遏 pL 的转录后不久, Cro 蛋白浓度增到一定程度, 使得操纵子 oR 也被占据, 从而阻断了由右向启动子 pR 处开始的转录子 R3 的合成。由于有足够量的 R3 所编码的蛋白质, 所以终止是很完全的。

5B. 从转录子 R3 翻译出基因 Q 蛋白产物和晚期 mRNA 的合成。从感染时刻起, 微量的转录子 R4 就已开始连续不断地合成, 它从右向启动子 pR2 处开始, 到终止位点 tR4 处停止。这个转录子不编码任何已知的基因产物。但为晚期 mRNA 编一个前导序列 (leader)。一旦转录子 R3 开始形成(见 3), 立即就生成基因 Q 蛋白的产物。Q 蛋白关系到晚期 mRNA 的合成, 而晚期 mRNA 又编码结构蛋白质、装配蛋白、成熟体系以及裂解酶。Q 蛋白也是一个抗终止因子, 它作用于 qut 位点, 使大肠杆菌的 RNA 聚合酶不在 tR4 处终止转录, 因此转录子 R4 延伸形成转录子 R5, 即晚期 mRNA。



## 晚期转录的时间顺序

由于 O、P 和 Q 基因的产物是由同一个转录子 (R3) 产生的, 很重要的一点就是在产生大量的 DNA 复制蛋白之前, 晚期蛋白 (例如构成 DNA 包装体系的一些蛋白质), 是不会合成出来的。调控因子将 DNA 复制的时间与晚期 mRNA 翻译产生活化因子的时间区分开来。这些调节因子有以下四个特点:

1. 某些 O、P 基因的产物是从转录子 R2 转译而得, 使 DNA 复制有一个小的头部起点。

2. Q 基因产物的形成及 Q 蛋白积累到足以发生抗终止作用的浓度是需要一定时间的; 在此期间之内, 领先发生的是复制作用。

3. 转录子 R5 的合成被基因 cII 蛋白所延滞 (途径不明)。

4. 晚期转录子很长并含有许多顺反子, 因此, 需要几分钟来完成其合成。基因处于下游时, 它们的翻译过程比上游的晚几分钟开始。最后, 装配头部的和切断  $\lambda$  DNA 进行包装的体系都十分复杂, 这些过程都是很耗时的; 另外, 裂解酶效率极低, 需达到足够高的浓度才能使裂解发生。

总之, 需要大量的晚期 mRNA 所编码的结构蛋白质。因此, 若需使晚期 mRNA 合成高速度进行, 就必须解除 cII 蛋白对 Q 蛋白介导的抗终止作用的抑制。这可以由 Cro 蛋白通过关闭转录子 R3 的合成作用来实现 (见前面 5A)。R3 编码 cII 蛋白。由于 mRNA 是不稳定的, 在几分钟之内就不再有供 cII 蛋白合成的模板存在。这个蛋白质是短寿命的。一旦停止合成转录子 R3, cII 蛋白的浓度就迅速降低。于是, Q 蛋白参与的抗终止作用不再受抑制, 晚期 mRNA 合成就开始 (实际上, cII 蛋白的延迟对 cII 突变株形成噬菌斑并不重要。然而, 这些突变株的裂解比野生型噬菌体更快, 而且只产生约为噬菌体正常数目一半的子代)。

产生大量晚期 mRNA 需要进行 DNA 的复制, 但与噬菌体 T4 的情况相反, 这仅仅是个“基因剂量” (gene dosage) 效应, 一旦形成了 O、P 基因的蛋白产物, DNA 的复制就开始。复制作用不仅为新噬菌体准备子代 DNA 分子, 而且也扩增 RNA 聚合酶的作用模板——晚期基因。因此, 当头、尾和其他颗粒的组分开始合成时, 由于 DNA 模板数目的增加, 使晚期 mRNA 的合成速度 (用每分钟的分子数来测量) 也增加了。于是随着转录作用的增强, 就加快了噬菌体结构成分的合成速度, 从而加速噬菌体的装配速度。

## 裂解周期中 $\lambda$ 转录的小结

让我们回顾这个精致的调控体系的重要特点。很重要的一点是没有使用  $\lambda$  的专一性的 RNA 聚合酶; 而是在全部生活周期中一直使用大肠杆菌的 RNA 聚合酶 (同 T4 一样) 并且通过附加蛋白质对大肠杆菌 RNA 聚合酶进行修饰, 改变此聚合酶对于各种 DNA 碱基序列的专一性。然而, 在任何时候此酶对于启动子的活性也不被修饰, 只是它对某些特定的终止位点的识别, 即接受对转录的终止信号的能力受到破坏。转录作用被阻止是 Cro 蛋白发挥其阻遏活性的后果。而不是 Cro 蛋白对 RNA 聚合酶有什么直接的作用。通过标准的阻遏作用, 噬菌体不能再使用启动子 pL 和 pR。  $\lambda$  DNA 上特定

区域开始转录的时间是由下面几个需要时间的过程所决定的：(1) 对基因N及基因Q的蛋白产物的转录及翻译过程；(2) Cro 蛋白对 cII 蛋白合成的阻断过程；(3) 残存的 cII 蛋白的降解过程；Cro 蛋白的抑制活性避免了浪费性的合成。最后一点，由于一个巨大的 mRNA 分子编码所有的结构成分，这些成分还将被依次地转译，这个全成分的 mRNA 的合成也需要很多分钟。因而，完整的头部的合成及有功能的噬菌体成熟体系的合成，都推迟到 DNA 复制体系可提供出足够多的  $\lambda$  DNA 拷贝数时才能完成。延滞作用的结果，使噬菌体在成熟体系建立之前形成约 30 个拷贝的  $\lambda$  DNA，并使宿主细胞在它遭到裂解之前已形成约 100 个完整的噬菌体颗粒。

### DNA 复制和成熟：偶联过程

O、P 基因的产物合成后，环状的  $\lambda$  DNA 开始复制， $\lambda$  DNA 复制有两种模式，即  $\theta$  式复制和滚环式复制（图 15-31，也见第八章）。 $\theta$  模式增加了转录的和进一步复制的模板数。滚环模式提供出大量的、形成子代噬菌体的 DNA。当进行生活周期的循环时， $\theta$  式复制停止，滚环式复制成为主要形式。对于  $\theta$  模式已经了解得很多了。事实上，对于  $\theta$  式分子的分析为在所有的有机体内的复制叉的双向移动提供了第一个证据（图 15-31 见第八章）。复制起点的碱基序列已被测定。起点中有很高的 A + T 含量，可能有不寻常的结构。与复制起始作用有关的早期现象是在复制起点处发生单链断裂。

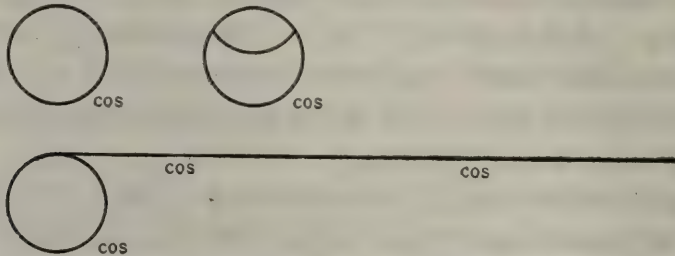


图 15-31  $\lambda$  DNA 在成熟开始时存在的三种形式。具有连接着的互补单链末端的区域（噬菌体头部中存在的线性 DNA 分子的末端）被称为 cos（即指粘性末端）。

有关滚环式复制作用的知识以及这种复制是如何起始的，目前知道的比  $\theta$  式的要少得多。由于  $\lambda$  DNA 的切割机理和前面已经讨论过的其他噬菌体的不同，因而对我们来说，这部分要讨论的最重要的一点，就是  $\lambda$  DNA 的复制作用与噬菌体成熟作用之间的关系。

在  $\lambda$  噬菌体颗粒中见到的 DNA 是线性的，有单链的末端的分子。当形成环状 DNA 时，这些单链末端相互连接，如此形成的双链区叫粘性位点（cohesive site）。所以，每一个单体的  $\lambda$  环含有一个 cos 位点，而一个多体的滚环分支则含有许多个 cos 位点。

在  $\lambda$  噬菌体颗粒中，DNA 分子的末端总是单链的末端，所以一定存在着一种机理来对 cos 位点切割，产生出这些单链末端。有两种可能的切割方式：按位点专一性的方式切割和按测量长度进行切割的两种体系。因为切割精确性的要求太苛刻（测量一个具有 50 000 碱基对的 DNA 分子，要求精确到 1 个碱基，所以用第二种方式立即去识别出切割点进行切割，看来是不可能的。又通过观察  $\lambda$  突变体的不同长度的 DNA 的包装效



率,也认为第二种方式是不可能的,因为只有 DNA 长度不小于正常  $\lambda$  DNA 长度 73%,或不大于正常  $\lambda$  DNA 长度 107%,并且其单链末端为与正常大小的 DNA 的末端的碱基序列相同的, DNA 才能够被有效地包装到噬菌体头部。因此,似乎位点-专一性的切割体系是更可能存在的方式,这种体系被称之为终止酶 (terminase) 或终止体系 (ter system),这种 DNA-切割就被称为终端切割 (ter-cutting)。

图 15-31 示出了三种主要的细胞内的  $\lambda$  DNA-环状分子、 $\theta$  式分子及滚环式分子。随着时间发展,噬菌体的头部及尾部已经合成, Ter 体系被活化,滚环式复制就成为主要的方式。因为滚环有两类粘性位点——一类在环中,另一类在线性分支中,必然地存在着某些方式以避免切开在环中的那种粘性位点。如果环中的粘性位点被打断,那么复制作用就要被中止。这个难题通过 Ter 体系对位点的要求而克服了,即只有单个 cos 位点时并不发生有效的切割;但如果存在着两个 cos 位点,而且它们都在同一个 DNA 片段上,那么切割就可能发生。切割两个相邻的粘性位点,可以从  $\lambda$  DNA 的线性分支上切下一个  $\lambda$  DNA 分子;但是如果这一对切点,一个处于线性分支中,一个处于环中,则不可能切出一个  $\lambda$  DNA 分子。虽然对于形成一个切割并不需要  $\lambda$  DNA 分子一定是线性的,因为单个的  $\lambda$  单位也能从在遗传重组作用中已经形成的二聚体的环状分子上切下来,但是在单体的环状分子中决不会发生切割。“两个 cos 位点”的规则解释了为什么第 1 个  $\lambda$  DNA 单位一定是从滚环的串联分支上切下来的,然而,这个规则并没有允许切除第 2 个 (即相邻的那个)  $\lambda$  单位。因为第 2 个单位只有一头有一个粘性位点,另一头是一个游离的粘性末端,因此这份 DNA 只有一半是可用的,因为只有交错着的 DNA 片段才能被包装 (见图 15-32)。这个过程是不经济的,其实,游离的粘性末端再加上相邻的 cos 位点也足够发生 DNA 切割,并且这样的切割还可以顺序地进行包装。因此,可能存在着精确的终端切割规则。

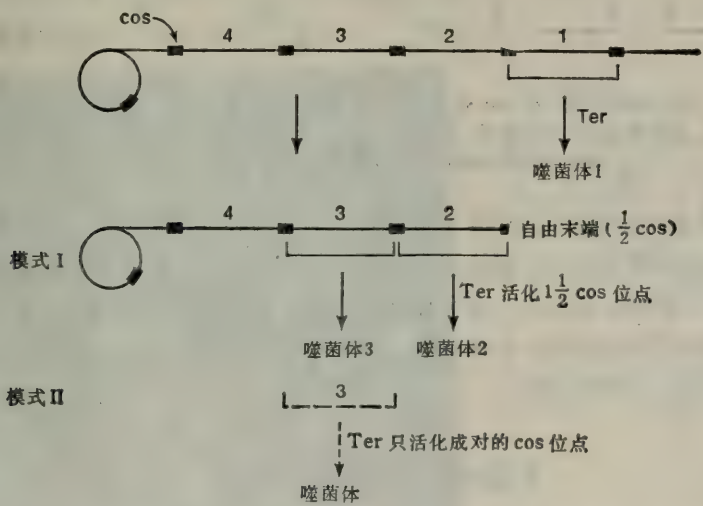


图 15-32 包装的两种规则: 第 I 种模式中每一个  $\lambda$  单位都被包装。第 II 种模式是较受限制的, 带有交错末端片段的单位 (3) 才被包装。其中第一种模式较为经济,  $\lambda$  噬菌体使用的是第一种。

图 15-33 中给出的一个精巧的实验, 示出游离的末端必然是接近于 A 基因的一端 (见

图 15-29)\*。在这个实验中,先制备三种类型的  $\lambda$  DNA 分子,即用体外的粘性末端连接法,将或者是两个完整的 DNA 分子连接 (I), 或者是一个完整的 DNA 分子连上一段含有基因 A 粘性末端的 DNA 片段 (II) 或者是一个完整的 DNA 分子连上一段含有基因 R 粘性末端的片段 (III)(这类片段的制备法是: 先将完整的 DNA 分子在接近中央处切断,然后将带有特定末端的片段与带着另一种特定末端的片段分离,从而获得所需要的片段)。这些分子在 P 基因处都有一个突变,所以不可能发生 DNA 复制。将具有这类 DNA 的噬菌体分别地去感染三份大肠杆菌的细菌培养液。在感染中不会发生 DNA 复制作用,但是有转录作用,结果合成出头、尾及 Ter 体系的成分。如果 Ter 体系可以从这些杂交分子上切下  $\lambda$  DNA 单位,那么就会产生噬菌体颗粒,其结果如图所示。当大肠杆菌细胞受二聚体分子感染,或受带有游离 A 基因的分子感染,都能产生噬菌体;而用仅仅带有游离 R 末端的杂交分子去转染时,不产生噬菌体。这个实验以及其他的实验指出  $\lambda$  DNA 从滚环的线性串联分支分子上切下的包装过程是有极性的,是先从 A 端起,再到 R 端的。

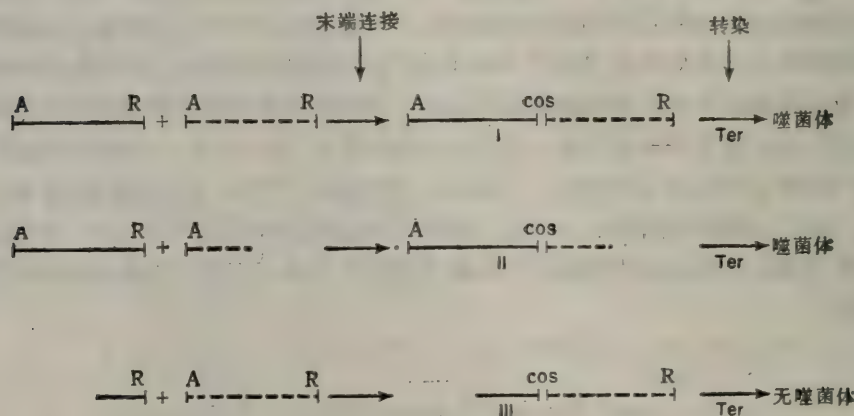


图 15-33 这个实验表明一个 cos 位点和一个游离的 A 基因的粘性末端对于 Ter 系统的活性是足够的。粘性末端用小竖线表示。所有的分子都在  $\lambda$  的 P 基因中有一个突变,以避免 DNA 复制。转染指细菌的原生质体受  $\lambda$  噬菌体 DNA 的感染。

在 cos 位点的切割和  $\lambda$  DNA 的包装以某种方式相偶联,事实上如果 Ter 蛋白质不是空的  $\lambda$  头部的组成成分,那么在一般的情况下 Ter 体系是无活性的。因此,当细菌被一个不能形成完整头部的  $\lambda$  突变体感染时(例如  $E^-$  突变体就不能制造主要的头部蛋白),DNA 滚动环的线性分支就不被切断。在这样的突变体系中,具有两个 cos 位点的二聚体同样地也不能被切断。

### 大肠杆菌噬菌体 $\phi$ X174

以下多方面的原因表明对于大肠杆菌  $\phi$ X174 的研究是非常重要的。 $\phi$ X174 是人们

\* A 基因及 R 基因并不重要,我们用这些基因的符号给  $\lambda$  DNA 的两端定名,以避免在称呼左边及右边时混淆。



发现的第一个以单链环状 DNA 作为遗传物质的有机体;该 DNA 是第一个能在体外复制产生具有生物活性的 DNA,并且通过对该 DNA 复制方式的研究产生了 DNA 滚环复制假说;这种噬菌体非常小,只有少数几个基因,而且是相当简单的(表 15-3);对  $\phi X174$  噬菌体 DNA 的碱基序列的研究以及对其蛋白质的氨基酸序列的研究首次指出基因能够按重叠的阅读框架被翻译等。

大肠杆菌噬菌体  $\phi X174$  是一个环状、单链的 DNA 分子(图 15-34)。

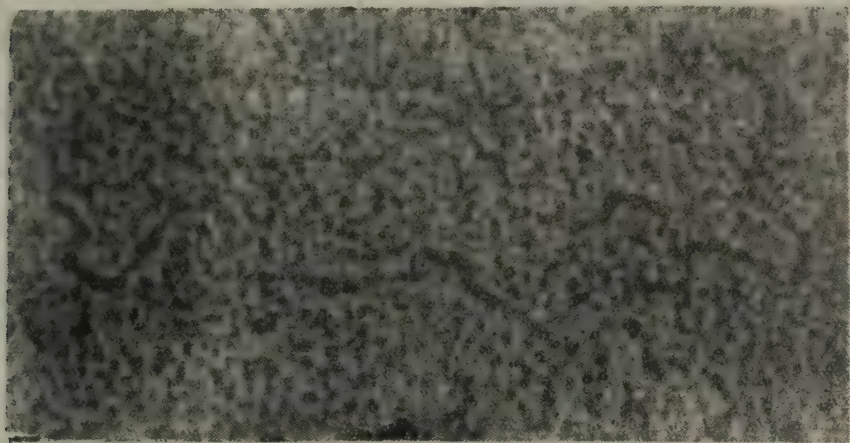


图 15-34 环状单链  $\phi X174$  DNA 分子。此照片显示的是用电子显微镜所看到的最初的单链 DNA 分子(承 Albrecht kleinschmidt 提供)。

它具有 5 386 个核苷酸,并且已知其碱基序列, 这个 DNA 分子被包在一个由三种外壳蛋白及一种内部蛋白组成的 20 面体的头部中(图 15—35)。

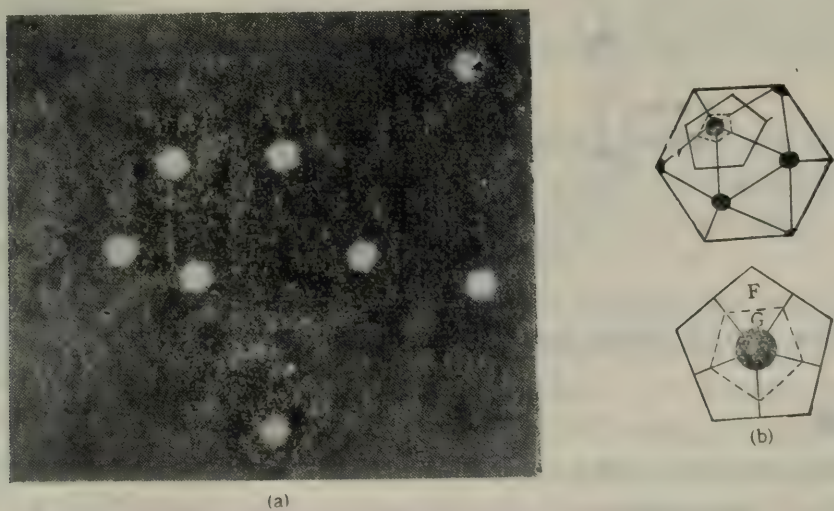


图 15-35 噬菌体  $\phi X174$ 。(a) 电子显微镜图(承 Robley williams 提供)。(b) 噬菌体  $\phi X174$  颗粒中亚基(上图)和放大的穗状区域(下图)的示意图。表明了基因 H、基因 G 和主要的外壳蛋白的基因 F 的定位排列。

重叠基因

表 15-7 列出了噬菌体  $\phi$ X174 所具有的 11 个基因，而且每个基因的产物都已被分离出来，这些蛋白质所含有的氨基酸总数超出了  $\phi$ X174 这个很小的 DNA 分子的编码容量。但是， $\phi$ X174 已经有效地进化，产生一些可以用不同的阅读框架进行翻译的重叠基因，如图 15-36 中的基因图所示。例如基因 B 和 E 分别被包含在基因 A 和基因 D 中。重叠阅读框架是如何进行翻译的将在第十二章进一步介绍。

表 15-7 噬菌体  $\phi$ X174 基因和功能

基 因	功 能
A	RF 复制;病毒链合成
A*	关闭宿主的 DNA 合成
B	形成衣壳(capsid)
C	形成噬菌体的单位长度的 DNA
D	形成外壳的
E	裂解细菌
F	主要的外壳蛋白 (coat protein)
G	主要的突穗状蛋白 (spike protein)
H	次要的突穗状蛋白;对宿主的吸附
J	核心蛋白,子代DNA进入噬菌体颗粒
K	未知

A\* 见图 15-36。

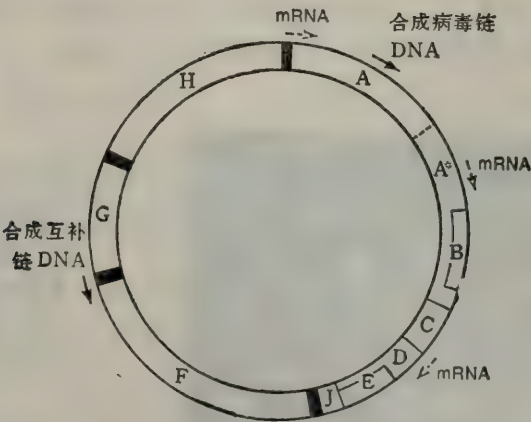


图15-36  $\phi$ X174 的遗传图谱。表示噬菌体基因以及复制和转录起始位点,间隔区为粗黑色部分。

吸附和穿透

与那些注射时才把蛋白质注入的 T 系列的噬菌体不同，噬菌体  $\phi$ X174 通过突穗状蛋白 (spike protein) 先吸附到大肠杆菌上(图 15-35)。 $\phi$ X174 的末端突穗状蛋白 (基因 H 和 G 的产物)和  $\phi$ X174 DNA 一起进入细菌中，它(基因 H 产物)以一种未知的方式发生作用，使细菌转变为噬菌体工厂。人们已经观察到一些被称为导航蛋白 (pilot protein)



的类似蛋白质在许多很小的噬菌体里。这些小的噬菌体不具有关闭宿主大分子合成的机理,因此人们假设这些蛋白质通过某种方式把细菌 DNA 聚合酶引导到噬菌体 DNA 上,指导噬菌体 DNA 进到宿主 DNA 复制装置(replication apparatus)上。虽然这种功能还没有实验证明,但是人们认为这种功能对于噬菌体 DNA 的复制速度和效率是必要的。通过这样的速度和效率,小的噬菌体 DNA 分子才得以和大大过量的细菌 DNA 分子进行竞争。

$\phi$ X174 DNA 不注入细菌,而是在细胞表面释放噬菌体 DNA。因为当噬菌体颗粒吸附到细菌上的短时间内,如果将 DNA 酶加入到这个细菌-噬菌体的悬浮培养液(suspending medium)中,噬菌体 DNA 将被降解(对于任何有尾的噬菌体来说,其 DNA 并不被 DNA 酶消化)。外壳蛋白的除去和亲代DNA的复制是互相偶联的。关于 DNA 外壳蛋白除去、DNA 分子转移到细胞内以及转移过程中的复制作用等的机理目前尚不了解。

$\phi$ X174 DNA 的复制

为了讨论单链 DNA 分子的复制,将使用下面这些专有名词。病毒颗粒中含有的以及任何具有相同碱基序列的 DNA 链都称为(+)链;具有与其碱基序列互补的 DNA 链称为(-)链。RNA 是从 DNA (-)链转录的。

$\phi$ X174 DNA 的复制包括以下几步(表 15-8):

表 15-8  $\phi$ X174的复制环

步 骤	时间(分)(在33℃)	事 件
SS→RF	0—1	吸附和穿透;病毒单链转变为亲代的 RF; RF 转录
RF→RF	1—20 25	亲代 RF 复制,形成 60 个子代 RF; RF 分子及宿主 DNA 的复制停止
RF→SS	20—33 40	大约从子代 RF 产生 35 个滚环;并由此产生大约 500 个 SS 分子; DNA 被插入到噬菌体颗粒中发生裂解过程

注: SS 代表单链噬菌体 DNA, RF 代表环状复制型。

1. 由亲代单链 DNA 分子(病毒的或(+)链DNA)转变成共价、闭环的双链分子,这种分子称为复制型 I(RFI)。由于此步发生在转录开始之前,所以仅仅依靠宿主的酶。新生链即(-)链是编码链,只有此链才能够被转录。由(+)链转化为 RFI 的酶学机理将在第八章详细论述。

2. 合成 RFI 的多个拷贝。转录(-)链,并产生一个多功能的蛋白质,即基因 A 蛋白。A 产物在(+)链上的 4 297 和 4 298 碱基之间产生一个缺刻(此处称为复制起始点, ori), A 蛋白本身则共价地连接在 5'-P 末端,如图 15-37 所示。这个有缺刻的 DNA 分子被称为 RFII。然后,在侵染之前就已合成出来的一些大肠杆菌蛋白质引起 RFII 亲代(+)链 DNA 分子的一部分被新加入的(+)链取代。这个取代是通过带泡的滚动环复制方式(looped rolling circle replication)完成的(细节已在第八章讨论过)。当完成复制的一个循环时,这条被取代的(+)链又从带泡的滚动环上被切下来重新环化,并作为合成另一条(-)链的模板起作用,结果产生一个新的 RFI 分子。

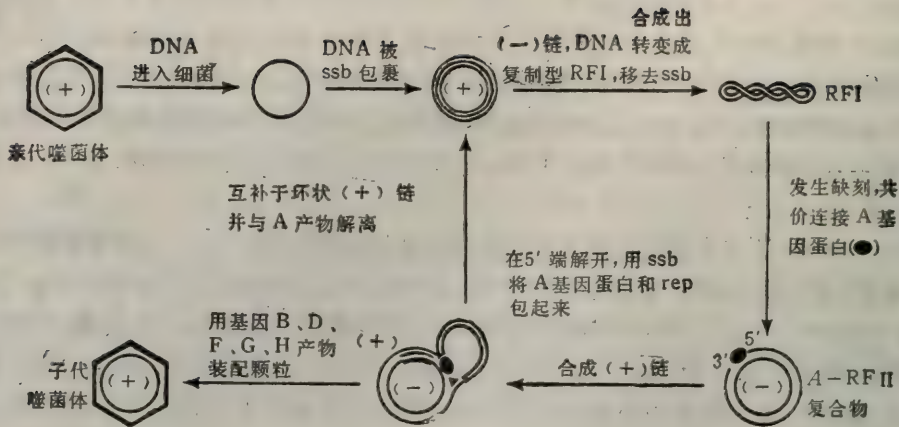


图 15-37  $\phi$ X174 DNA 复制的两个阶段。首先亲代噬菌体 DNA 转变为双链超螺旋 RFI。然后,基因 A 蛋白在其上制造一个缺口,滚动环复制开始,进而在形成 RFI 的过程中复制出来的子代(+)链取代 RFI 中的一部分。在有噬菌体编码的成熟蛋白和外壳蛋白同时存在时,子代的(+)链环被包装,而不是转变成 RFI 分子。除了噬菌体的基因 A 蛋白,所有复制过程中所使用的蛋白质都是细菌的蛋白质。

### 3. 用于包装的(+)链的合成。

在合成 RFI 过程转变为合成自由的(+)链的过程的控制中,不存在什么开关。相反,在子代(+)链作为模板合成新的(-)链之前,包装体系就已捕获子代(+)链。这一捕获过程要等到完整的噬菌体头部几乎已合成完毕,并且大约已有 35 个拷贝的 RFI 时才开始。此时,大部分 DNA 还在进行滚动环复制,产生新的 RFI 分子和一些用于包装的(+)链,宿主 DNA 的合成这时已完全停止。造成这种结果的部分原因是由于 A 蛋白的一种未知功能;另外的原因是所有的宿主的 DNA 聚合酶都被用来合成  $\phi$ X174 DNA 了。当更多的头部被合成时,所有被取代下来的(+)链都被包装起来,由于缺乏模板,RFI 的合成也随之停止了。包装机理将在后面的章节中讨论。

### $\phi$ X174 DNA 的转录

$\phi$ X174 噬菌体具有 3 个启动子,并且这三个启动子同时被激活,除对溶菌酶外,不需要任何短暂性调节。所有的转录都是以(-)链为模板进行的,因此只有当最初的(-)链被合成时,才能开始转录过程;从这时起,RFI 的合成一直继续下去,直到形成足够量的头部蛋白并开始包装,RFI 的合成才停止。由于  $\phi$ X174 颗粒很小,在头部蛋白合成出来并装配成头部之前, $\phi$ X174 已完成了很多次的复制过程(大约每两秒钟一个复制周期),在感染 24 分钟后出现第一个完整的  $\phi$ X174 颗粒。由于翻译过程较慢,而且溶菌酶不是很活跃,所以溶菌酶的活性被延滞,因此,直到感染 40 分钟之后裂解才会发生,此时大约合成出 500 个噬菌体颗粒。

### $\phi$ X174 DNA 的包装

$\phi$ X174 DNA 的包装需要 7 种噬菌体蛋白质,其中 4 种存在于完整的颗粒中。装配



的步骤如下(如图 15-38 所示):

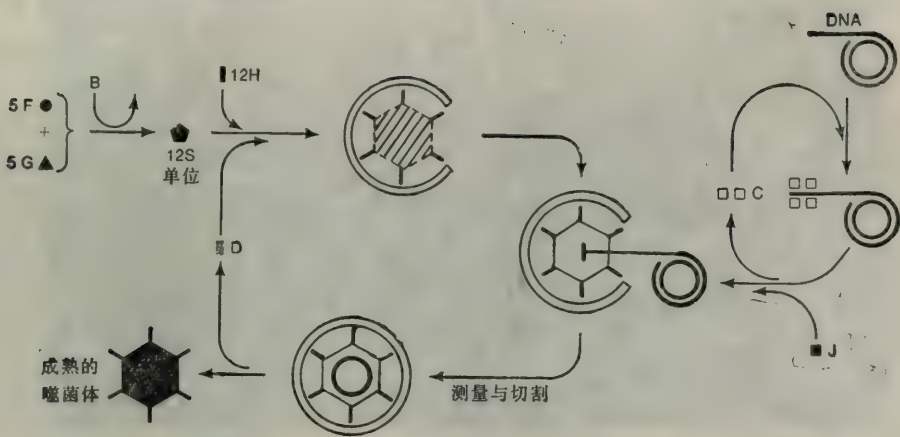


图 15-38  $\phi$ X174 噬菌体装配模型。字母代表  $\phi$ X174 蛋白。B 蛋白催化 5 个 F 蛋白分子(主要的头部蛋白)与 5 个 G 蛋白分子(穗状蛋白)的聚合,形成一个 12S 单位。D 蛋白在形成 108S 颗粒过程中可能具有支架的(scaffolding)功能。C 蛋白加速  $\phi$ X174 DNA 分子的包装。A 蛋白测量 DNA 的单位长度,并且使之形成环状。对 J 蛋白则了解甚少。

1. 头部蛋白(基因 F)和穗状蛋白(基因 G)形成五聚体。在基因 B 蛋白的影响下,该五聚体进一步形成称为 12S 颗粒的聚集体。
2. 基因 H 蛋白参与突穗(spike)的形成过程。
3. 基因 D 蛋白形成一个框架,在这个框架上,由 12S 单位形成头部前体。
4. 图中的基因 C 蛋白结合在取代的(+)链 5' 端上,在一定程度上引导 5' 端进入头部前体。基因 J 蛋白很可能最初结合在头部前体的 DNA 外面,以一种未知的方式编排头内的 DNA,并巩固所形成的结构。
5. 基因 A 蛋白在滚动环 DNA 上切割取代的(+)链,形成泡状的滚动环,并环化(+)链。
6. 基因 D 蛋白被移走(很可能是自动地在头部关闭之后被移走),从而形成完整的噬菌体颗粒。

目前已可在体外合成亲代的第一个 RFI 分子和子代的 RFI 分子,并能进行滚动环复制。同时,如果存在着 7 种有关形态发生的蛋白质,那么,可以同时发生包装反应和滚动环复制。既然已存在着偶联的转录-翻译体外体系,那么应当可能从 RFI 合成出所有的噬菌体蛋白。当把这些体系联合到一起时,  $\phi$ X174 将可能是第一个完全由已确定的分子所合成的完整的生物有机体。

### 丝状 DNA 噬菌体

已经系统地研究了三种丝状大肠杆菌噬菌体: M13、fd 和 f1 (图 15-39)。这几种噬菌体都有一个环状的单链 DNA 分子。它们非常相似,通过研究其各自的情况并进行综合,确定了各个噬菌体的生活周期。由于这些噬菌体的生命周期与  $\phi$ X174 有许多相似之处,所以不作进一步介绍,只讨论噬菌体 DNA 的穿透作用和完整的噬菌体颗粒的释

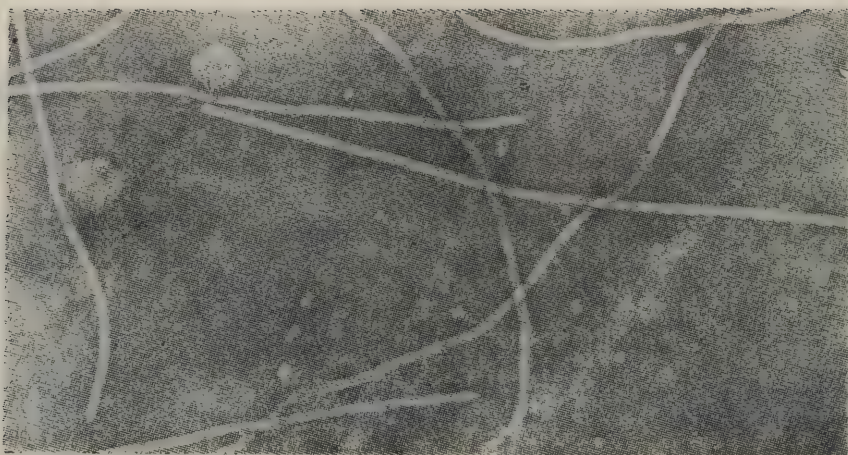


图 15-39 丝状大肠杆菌噬菌体 M13 的电子显微镜照片。大的和小的球状颗粒是细胞碎屑 (承 Robley Williams 提供)。

放。

在所有已知的烈性的噬菌体中，只有丝状大肠杆菌噬菌体既不杀死宿主也不引起裂解。丝状噬菌体吸附到 F 菌毛的顶端上。F 以一种未知的方式 (已有许多理论讨论

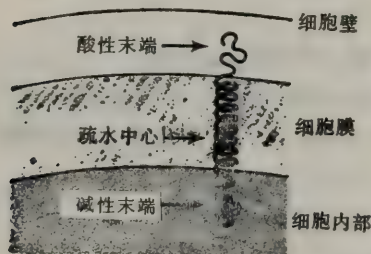


图 15-40 M13 外壳蛋白在细胞膜上定位的示意图。

这种方式), 将噬菌体颗粒带到该细菌的菌毛基部的细胞壁上。一种理论是颗粒沿着菌毛一侧的沟槽下行进入细胞, 另一种理论是菌毛通过吸收菌毛蛋白质到细菌内而发生作用 (这是用 pili 所观察到的现象)。可以肯定的是, 整个噬菌体以完整的形态穿透细胞壁并附着在内层细胞膜上。这时, 噬菌体颗粒除去外壳, 同时单链 DNA 转变为 RFI 型。这些过程和  $\phi X174$  相同, 都是偶联的。接着外壳被降解, 蛋白质的亚基在子代颗粒中可以再次被使用。这几种噬菌体的外壳蛋白

质能透入细胞是非常独特的。

丝状噬菌体复制周期的细节尚不很清楚。它像  $\phi X174$  一样有一个通过滚动环机理形成 RFI 的阶段。打开 (+) 链合成的开关需要基因 5 蛋白, 这种蛋白质是一种单链 DNA 结合蛋白, 它包裹着滚动环的尾部, 这是与  $\phi X174$  不同之处。而且, 这个开关并不与包装作用相偶联, 而是那个被基因 5 蛋白包裹着的线性 (+) 链继续地在滚动环上被切割。然后以一种未知的方式环化。因此在细胞内存在着大量的 (+) 链和基因 5 蛋白的复合体。

主要的包装蛋白即基因 8 蛋白大量地合成并贮存在细胞内膜上。在细胞质内没有自由存在形式的基因 8 蛋白。这种蛋白质是一个具有三个不同区域的  $\alpha$  螺旋: 一个碱性末端、一个疏水中心 (hydrophobic center) 和一个酸性末端 (图 15-40)。人们假设 (基于很少的数据) 疏水中心使得该蛋白质能够定位在疏水膜上, 酸性末端朝着细胞壁, 碱性末端朝着细胞内部。它很可能是代替基因 5 蛋白, 并结合于 (+) 链上。这个还含有其他两种蛋



白质的噬菌体颗粒以一种很不清楚的方式在细胞内膜上装配。对每一个完整的颗粒来说，有一个芽在细胞表面上形成，就会挤出一个噬菌体（图 15-41）。这一过程可以无限地进行下去，而对细胞没有损害。

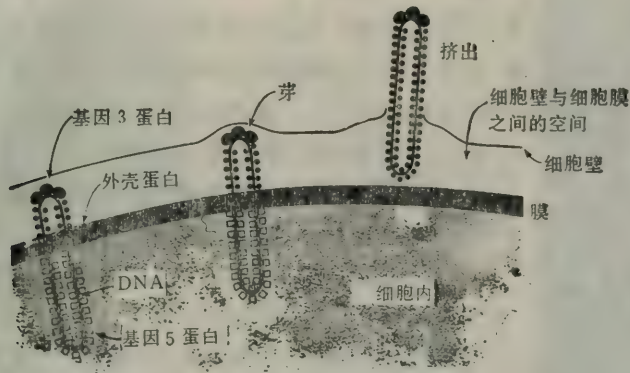


图 15-41 M13 噬菌体的装配和分泌。基因 5 蛋白覆盖着 DNA，在一定程度上使得该 DNA 分子能穿透细胞膜。在细胞膜中外壳蛋白分子代替基因 5 蛋白结合到 DNA 分子上。基因 3 蛋白附着在该颗粒的一端并参与吸附作用和分泌作用。

对典型的噬菌体来讲，噬菌斑是细菌裂解的结果。然而，丝状噬菌体形成噬菌斑而不裂解宿主细胞。原因是受感染的细胞仍能继续地生长和分化，但比未受感染的细胞要慢得多，推测可能是因为宿主的复制机器在宿主细胞生活周期的很大部分中被用来制造噬菌体 DNA 了。这种缓慢生长足以在细菌的菌苔上产生一个较混浊的区域，这就是丝状噬菌体的不十分清亮的噬菌斑。这些噬菌斑只有在适当的时候才能观察到，因为菌苔上未被感染的正常细菌由于营养耗尽而停止在培养基上继续生长时，那些被感染的细胞则在噬菌斑处还可以继续地生长大约一个世代以上的时间，从而增加噬菌斑处的浊度，直到噬菌斑消失。

## 单链 RNA 噬菌体

已分离到几种单链 RNA 的噬菌体，了解得最清楚的是大肠杆菌噬菌体 f2、R17、MS-2 和 Q $\beta$ 。每一种都像  $\phi$ X174 一样，是没有尾的 20 面体（图 15-42）。MS-2 的遗传图谱如图 15-43 所示。通过对每种噬菌体的不同性质进行比较已经阐明了它们的很简单的生活周期。其 RNA 是具有大量分子内氢键的线性分子。由大约 3 600 个核苷酸组成（已知 R17、MS-2 和 Q $\beta$  的序列），具有 3 个基因，分别编码外壳蛋白（CP）、吸附蛋白（A）和 RNA 复制酶（Rep）。此 RNA 分子既是复制模板，又是 mRNA。因此，（1）在生活周期中既不再需要 DNA 聚合酶，也不再需要 RNA 聚合酶；（2）调节作用必然发生在翻译水平上。典型的噬菌体爆破值是 5 000—10 000。这个数值和典型的 DNA 噬菌体爆破值相比是非常大的（DNA 噬菌体的爆破值是一至几百）。这些颗粒在每个细菌中形成巨大的结晶体结构（图 15-44）。这些结晶体在一定程度上损害细胞膜，可以在缺乏溶菌酶的情况下引起细胞裂解。因此，与其他噬菌体不同，它的裂解时间是不严格的，在感染 30—60 分钟后细菌的裂解逐步地发生。

图 15-45 是 RNA 噬菌体的生活周期示意图,分阶段介绍如下:

1. 进入细胞并结合到核糖体上。在进入细胞内后,一个核糖体就立即结合到 CP 基因起始区的结合位点上。这个基因是 RNA 分子中最中心的基因。

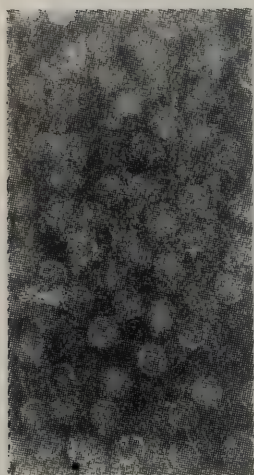


图 15-42 噬菌体 MS-2 的电子显微镜照片。注意少数几个噬菌体颗粒是中空的。

2. 翻译作用和三种蛋白中每一种的量的调节作用。如同第十三章中的解释的那样, RNA 的二级结构堵塞了 R 17 中的 A 蛋白和 Rep 蛋白以及 Rep 蛋白和核糖体的结合位点。然而, CP 基因的翻译却又打开了 Rep 蛋白的结合位点。因此,开始两种蛋白质都形成(指 CP 蛋白和 Rep 蛋白),但是后来由于结合到 rep 位点上的 CP 蛋白量不断地增加,也就抑制 rep 基因的翻译,减少 Rep 蛋白的量(见第十四章)。这种抑制作用是符合效率要求的;作为 10 000 个噬菌体的结构成分大约需要  $2 \times 10^6$  个拷贝的 CP 蛋白,而复制酶(即 rep 蛋白)仅仅需要催化量即可。A 蛋白的合成将以后再作简单的讨论。

3. RNA 噬菌体的复制(图 15-46)。只详细研究了 Q $\beta$  复制酶。此酶是一个四聚体,由 1 个 Rep 蛋白分子和 3 个宿主蛋白分子,蛋白质合成过程中的 EF-Ts、EF-Tu 翻译延长因子和核糖体蛋白 S1 (S1 的功能不清楚)所组成。复制酶以病毒的(+)链为模板,产生出(-)链。在进行合成时,此(-)链仅仅只在聚合作用的位点与(+)链接触,因此,大部分复制产物是单链的。在第一个(-)链的合成完成之前,另有几个(-)链开始合成,而且复制的形式是分叉状的。复制释放出的(-)链立即被复制酶用来作为模板合成(+)链,一些(+)链回到核糖体上去合成更多的 CP 蛋白,另外一些(+)链则被包装。所有的子代噬菌体中都只含有(+)链。

4. A 蛋白的合成。由于碱基配对,在自由的(+)链上 A 蛋白的核糖体结合位点总是关闭的。但是,(-)链的复制则起始于紧接着 A 基因的 5'端。在(+)链刚刚开始合成之后不久,有一个 A 蛋白的核糖体结合位点处于自由状态的短暂阶段,这时(+)链的互补链还没有开始复制(图 15-47)。在此期间, A 基因开始被翻译。在上述结合位点关闭之前,这种翻译过程只可能有一、两个回合。因此 A 蛋白的数目和(+)链的数目大体上保持相等。由于每一个病毒颗粒只含有一个 A 蛋白分子,所以这种设计是很经济的,后来每一个 A 蛋白在一个未鉴定的结合位点上结合一个 RNA 分子。因此人们认为 A 蛋白促进 RNA 和 CP 分子的相互作用。A 蛋白保留在 RNA 分子上,并在随后的感染作用中进入

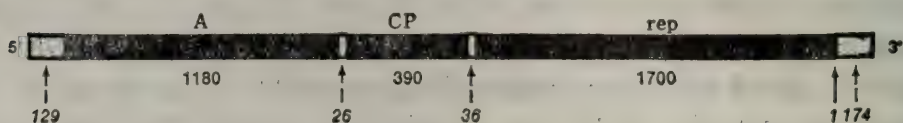


图 15-43 噬菌体 MS-2 的遗传图谱。表示每一个基因的所在位置和碱基数。斜体数字表示先导序列和间隔区的长度。



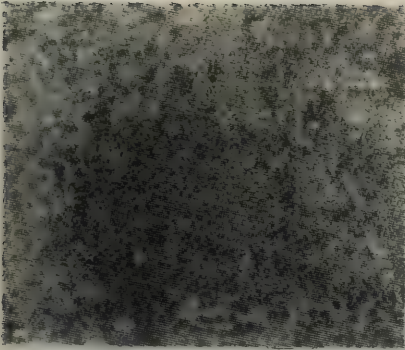


图 15-45 大肠杆菌中成熟的 MS-2 颗粒中的结晶结构。

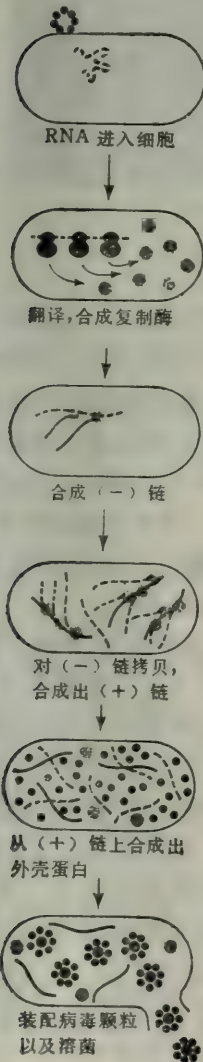


图 15-45 RNA 噬菌体生活周期示意图。

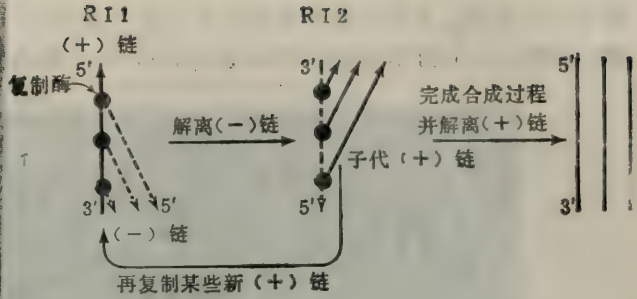


图 15-46 由 Q $\beta$  酶催化的 RNA 复制。R1 代表复制中间物。

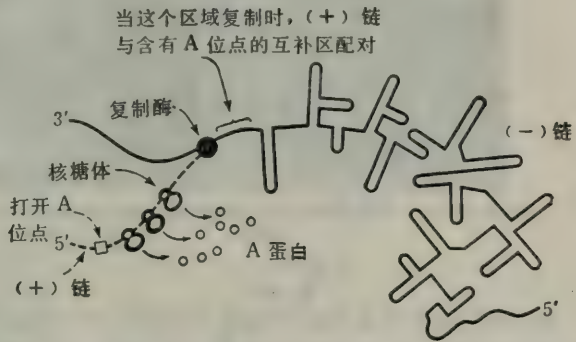


图 15-47 通过开放在新 (+) 链上的核糖体结合位点, 可以进行基因 A 蛋白的合成。

细胞。可能 A 蛋白的功能类似于  $\phi$ X174 噬菌体的穗状蛋白的功能。

5. 病毒颗粒的装配。一些 CP 蛋白分子同时自发地在一个新合成出来的 (+) 链周围聚集, 形成一个 20 面体外壳。

6. 细胞裂解。在大约有 10 000 个噬菌体颗粒形成之后发生裂解。但是没有产生噬菌体编码的裂解酶, 而且裂解机理尚不清楚。

## 一种双链 RNA 噬菌体: $\phi$ 6

已知的双链 RNA 噬菌体非常少。研究得最彻底的一个是  $\phi$ 6 噬菌体, 其宿主是假单胞菌 (*Pseudomonas phaseolica*)。这种噬菌体在结构上不同于大部分的细菌噬菌体。它的外壳蛋白被包裹在一个由蛋白质和脂类所组成的被膜之内 (同 15-48)。 $\phi$ 6 最特殊的性质

是具有三种双链 RNA 分子,分子量分别为  $2.3$ 、 $3.1$  和  $5.0 \times 10^6$ 。每一个 RNA 分子携带着不同的基因,人们说这种噬菌体带有分节基因组 (segmented, genome) (在第二十一章中将给出具有分节基因组的病毒例子)。这三种 RNA 分子共编码至少 12 种蛋白质。

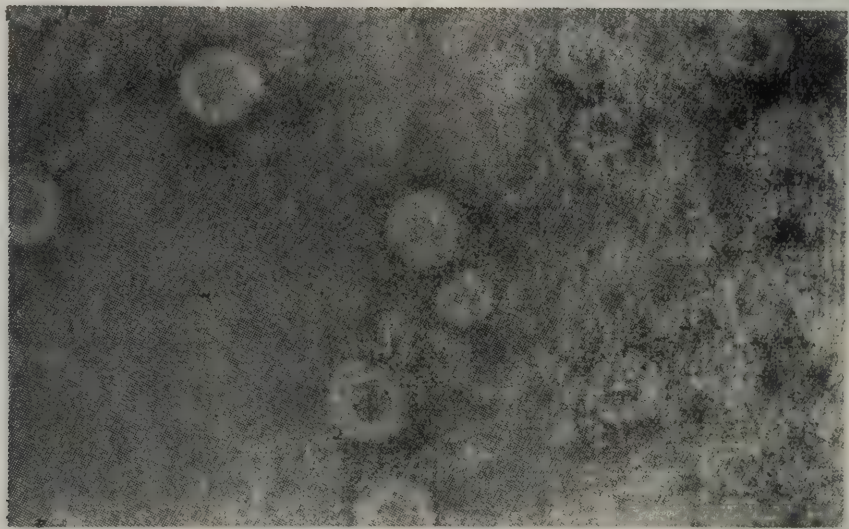


图 15-48  $\phi 6$  噬菌体的电子显微镜照片。较大的颗粒是完整的带被膜的颗粒、较小的颗粒缺少脂性被膜。

$\phi 6$  的复制和转录作用中有几个很有趣的问题。下面将仅仅讨论  $\phi 6$  噬菌体生活周期的有关特点。在适当的条件下细菌 RNA 聚合酶可在体外以单链 RNA 为模板合成出双链 RNA 产物。由于双链 RNA 对 RNA 合成来说并不是很好的模板,所以这个反应不能进一步进行下去,因为  $\phi 6$  的复制需要的不是细菌的 RNA 聚合酶,而是另外的酶。双链 RNA 也不能作为翻译的模板。如果把一个双链 RNA 注入到一个细菌中,它不能用细菌的酶进行翻译,因此也就不能产生噬菌体蛋白。在前一节中讲到的单链 RNA 噬菌体避免上述困难的手段是:首先对注入的单链 RNA 进行翻译,在翻译产物中有一个噬菌体编码的 RNA 复制酶,这个酶是可以用来产生子代 RNA 链的。

一个双链 RNA 噬菌体可以有两种选择:(1)或者是它必须解链,以产生一条编码链,这是一个在能量需求上完全不可能的过程;(2)或者它把一个能复制双链 RNA 的酶携带到细菌内。 $\phi 6$  噬菌体使用的是第二种机理,即噬菌体颗粒具有一个它所编码的有 RNA 聚合酶活性的物质,当噬菌体 RNA 进入细菌时,这个物质随之一同进入。这种活性物质尚未得到纯化,它对于产生噬菌体子代的蛋白外壳是必需的。人们估计这个物质是由噬菌体头部中几种不同蛋白质的区域接合到一起而产生的,其产生方式在很多方面相同于几个核糖体蛋白聚集成核糖体上的肽酰转移酶活性区域的方式(见第十三章)。

$\phi 6$  噬菌体 RNA 可以作为这种噬菌体的 RNA 聚合酶活性物质的底物 mRNA,分子可以通过链的取代反应产生(图 15-49)。噬菌体的 RNA 聚合酶活性催化从双链 RNA 的一端起始的合成,拷贝出非编码链,将编码链取代下来。被取代的编码链被作为 mRNA 起作用。第二个取代反应通常在第一个取代反应完成之前就开始了,如图所示。图 15-50 中的电子显微镜照片示出了一个正在进行这种取代反应的 RNA 分子和一



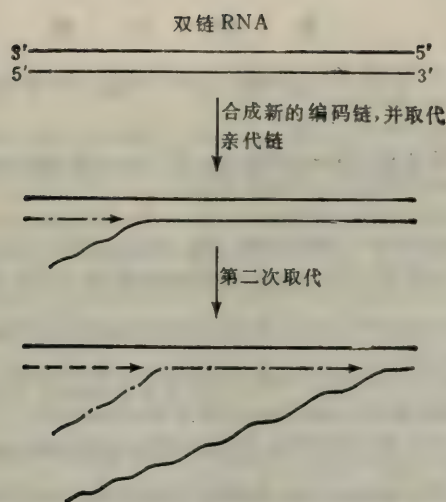


图 15-49 通过亲代 RNA 链的取代反应合成  $\phi 6$  的 mRNA。

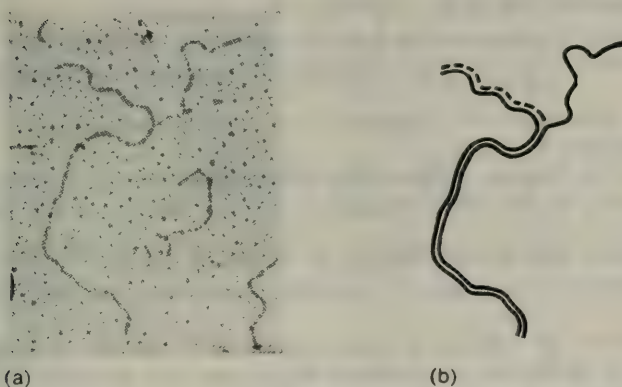


图 15-50 (a)  $\phi 6$  RNA 转录作用的电子显微镜照片。(b) 示意图。虚线表示正在合成的链。

个正在合成中的单链。

$\phi 6$  噬菌体的三种 RNA 分子中的每一种都可以通过取代反应作为转录的模板, 从这些 mRNA 可以翻译出至少 14 种蛋白质。

复制过程存在着一个特殊问题, 即取代反应产生子代单链而不能增加双链分子数目。现在通过实验已经排除了被取代的单链 RNA 再转变成双链子代 RNA 分子的可能性。目前虽然已经了解到  $\phi 6$  噬菌体复制作用和转录作用使用着同样一个酶活性物质。但是关于  $\phi 6$  DNA 复制的机理尚不清楚。

$\phi 6$  噬菌体子代噬菌体颗粒的产生需要的条件是:  $\phi 6$  噬菌体的三种双链 RNA 分子中的每一种的一个片段被包装在一个单个的噬菌体颗粒中。但是这个过程是如何发生的还不清楚。动物的呼肠病毒也同样都具有分节基因组。关于这个病毒的每种分节片段的各个片段被包装到噬菌体颗粒中去的方式将在第二十一章讨论。

## 参 考 文 献

- Adhya, S., S. Garges, and D. F. Ward. 1981. "Regulatory circuits of bacteriophage lambda." *Prog. Nucleic Acid Res. and Mol. Biol.*, 26, 103—118.
- Arber, W., and S. Linn. 1967. "DNA modification and restriction." *Ann. Rev. Biochem.*, 38, 467—500.
- Casjens, S., and J. King. 1975. "Virus assembly." *Ann. Rev. Biochem.*, 44, 555—611.
- Dubow, M. S. (ed). 1981. *Bacteriophage Assembly*. Alan Liss.
- Dunn, J. J., and F. W. Studier. 1981. "Nucleotide sequence for the genetic left end of bacteriophage T7 DNA to the beginning of gene 4" *J. Mol. Biol.*, 148, 303—330.
- Earnshaw, W. C., and S. R. Casjens. 1980. "DNA packaging by the double stranded DNA bacteriophages." *Cell*, 21, 319—331.
- Echols, H. 1980. "Bacteriophage development." In *Molecular Genetics of Development*. T. Leighton and W. F. Loomis. eds. Academic Press.
- Furth, M. E., J. L. Yates, and W. F. Dove. 1979. "Positive and negative control of bacteriophage lambda DNA replication." *Cold Spring Harb. Symp. Quant. Biol.*, 43, 147—153.
- Gottesman, M. E., S. Adhya, and A. Dar. 1980. "Transcription antitermination by bacteriophage lambda N gene product." *J. Mol. Biol.* 140, 57—75.
- Hershey, A. D. (ed.). 1971. *The Bacteriophage Lambda*. Cold Spring Harbor Laboratory.
- Kaiser, A. D., M. Syvanen, and T. Masuda. 1975. "DNA packaging steps in bacteriophage lambda head assembly." *J. Mol. Biol.*, 91, 175—186.
- Lewin, B. 1977. *Gene Expression. III. Plasmids and Phages*. Wiley-Interscience.
- Luria, S. E., J. E. Darnell, D. Baltimore, and A. Campbell. 1978. *General Virology*. Wiley.
- Simon, A., and R. Haselkorn. 1978. "Transcription and replication of bacteriophage  $\phi 6$  RNA." *Virology*, 89, 206—217.
- Rabussay, D., and E. P. Geiduschek. 1977. "Regulation of gene action in the development of lytic bacteriophage." *Comprehensive Virology*. 8. 1—196.
- Rosenberg, M., D. Court, H. Shimatake, C. Broad, and D. Wulff. 1980. "Structure and function of an intermediate regulatory region of bacteriophage lambda." In J. Miller, ed. *Operons*. Cold Spring Harbor Laboratory.
- Stent, G., and R. Calendar. 1978. *Molecular Genetics*. W. H. Freeman and Co.
- Studier, F. W. 1973. "Analysis of bacteriophage T7 early RNA and proteins on slab gels." *J. Mol. Biol.*, 79, 237—248.
- Szybalski, W. 1977. "Initiation and regulation of transcription and DNA replication in coliphage lambda." In J. C. Copeland and G. A. Marzluf. eds. *Regulatory Biology*. Ohio State University Press.
- Zinder, N. (ed.) 1975. *RNA Phages*. Cold Spring Harbor Laboratory.



## 第十六章 噬菌体 (II)

### 温和噬菌体及溶源生活周期

在上面所讨论的噬菌体的生活周期中,宿主细胞裂解,释放出许多子代噬菌体(最常见的效果),或者是不断地放出子代噬菌体而不杀死宿主(丝状噬菌体)。这一章中所讨论是另一种生活史周期,即溶源生活周期。在这种周期中,不产生噬菌体,宿主细胞能生存下来,并能无限地分裂。然而,宿主细菌在经历了很多世代以后,如果环境条件合适,噬菌体的溶源周期就会终止而开始一个新的裂解周期。一旦发生这种情况,宿主细胞被杀死,并释放出许多子代噬菌体,就像被噬菌体感染而起始的裂解周期一样。

能够进入溶源周期的噬菌体被称为温和噬菌体,它们有一个吸引人的方面,即它们也可以启动前一章所描述的裂解周期。细菌在受感染之前的生长条件和感染的倍数(multiplicity of infection)决定着噬菌体进入哪一条途径。至于最终走哪一条途径则是噬菌体的优良调控体系起作用的结果,这种体系不仅具有暂时性的调控作用,而且对环境条件十分敏感,例如对营养的供应,使得噬菌体有可能按不同途径生活。

### 溶 源 周 期

我们从讨论溶源性细菌的一般性质开始,然后探讨大肠杆菌  $\lambda$  噬菌体,它是研究得最清楚的温和噬菌体。

#### 溶源生活周期概述

有两种溶源周期,最普通的一种以  $\lambda$  噬菌体为典型代表,其生活周期如下(图 16-1)。

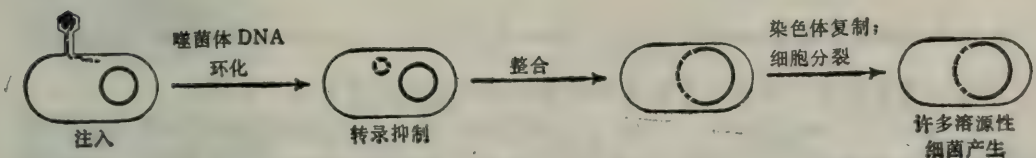


图 16-1 噬菌体 DNA 插到细菌染色体而完成溶源作用的一般模式。

1. DNA 分子注入到细菌中。

2. 经过一个短时间转录,合成出所需要的整合酶 (integration enzyme)后,阻遏物使转录终止。

3. 噬菌体 DNA 插到细菌 DNA 中。

4. 细菌继续生长和增殖,噬菌体 DNA 作为细菌染色体的一部分随之进行复制。

第二种较为少见,以大肠杆菌的 P1 噬菌体为其典型代表。与第一种不同之处是,第二种没有整合体,噬菌体 DNA 成为一个质粒(一个独立复制的环状 DNA 分子),而不是

宿主染色体的一部分。在这一章我们将主要讨论噬菌体的第一种生活周期。

### 一些有用的术语

下面的术语描述溶源周期的不同方面：

1. 既能进入裂解周期又能进入溶源周期的噬菌体，称为温和噬菌体 (temperate phage)。

2. 含有一套完整的噬菌体基因的细菌，称作溶源菌 (lysogen)。如果缺少了某些噬菌体基因，噬菌体在其中就不能完成裂解周期，这种细菌称作缺陷型溶源菌 (defective lysogen)。

3. 用温和噬菌体感染细菌形成溶源菌的过程叫做溶源化 (lysogenization)。

4. 如果噬菌体 DNA 进入了细菌 DNA，就说噬菌体 DNA 是整合 (integrated) 的 DNA。这个过程叫做整合或插入 (integration 或 insertion)，以质粒形式存在着的噬菌体 DNA 是不整合的。整合的或质粒状的噬菌体都称作前噬菌体 (prophage)。如果其中缺少了某些基因，这种前噬菌体就是缺陷型的。

### 溶源及溶源化作用的一般性质

溶源过程有两个重要性质：

1. 一种噬菌体感染细菌形成溶源菌之后，不能再被同种噬菌体感染，这种对“超感染”(superinfection)的抗性称作免疫性 (immunity)。

2. 在许多世代之后，溶源性噬菌体也能进入裂解周期，这个过程称作诱导 (induction)，噬菌体的基因作为一个 DNA 片段被剪切下来(图 16-2)。

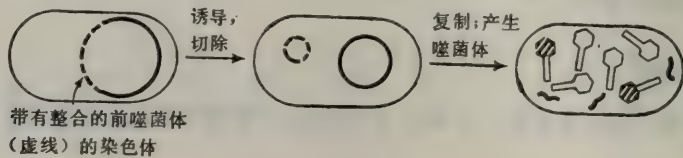


图 16-2 前噬菌体诱导作用的略图。前噬菌体 DNA 用虚线表示。右图中为了清楚删去了细菌 DNA。

很快将要讨论免疫性的分子机理及环境对诱导作用增强的分子机理。

在已知的几千种噬菌体中，90% 以上属温和性的。这些噬菌体通常不像 T4、T7 等高强度的烈性噬菌体那样，生成大量的子代，但可以通过另一种能力给予补偿——在不适于迅速产生子代的环境条件下，它们仍能扩增。当我们观察自然界中一个噬菌体可能遇到一群细菌时，就能明白上述的情况了。

我们先考虑正在活跃分裂的一大群细菌。如果一个噬菌体侵入了一个细菌并扩增 (在裂解周期中)，子代噬菌体的数目很快地增加。然而，由于周围的营养耗尽，细菌生长变得缓慢，在裂解周期中如果被感染的细菌停止生长，那么感染的噬菌体就会中途夭折 (噬菌体只能生长在代谢活跃的细菌中)。

当细菌缺乏营养时，在进入休眠状态之前，它们会降解自身的 mRNA 和蛋白质，待恢复供给营养后，细菌又重新生长。但是被噬菌体感染过的细菌 (其中噬菌体的生活周期



已被中断)就不是这样。一般来说,首先它永远失去了产生噬菌体的能力,这可能是因为宿主 mRNA 和蛋白质的降解破坏了严格平衡的噬菌体功能;接着细菌将会死去。如果细菌有溶源周期,也就是说噬菌体 DNA 有可能变为休眠状态,那么在这种溶源菌中的噬菌体就有可能避免受到破坏。当细菌生长得到恢复时,噬菌体基因作为细菌染色体的一部分而一起复制,即使这段时间噬菌体的产生暂时中断了,由于存在着诱导现象,所以仍然保留着产生噬菌体的潜在可能性。

现在让我们回到噬菌体对活跃生长着的细菌群体感染的问题上来,这时噬菌体在其中能快速地增殖扩增。当噬菌体数目超过细菌数目时,噬菌体就进入了它们的最后一个增殖周期;裂解以后,因为不再有可感染的细菌,噬菌体就不能再增殖。可能要好几年之后,它们才会再遇上敏感性宿主,在此期间,各种各样的破坏剂(如热、酸、蛋白酶、原生动物)都可能损坏噬菌体颗粒。事实上,噬菌体的吸附装置最易被化学试剂破坏。在再度遇到宿主之前,噬菌体是没有机会再增加它的数目的。但是,在大剂量感染(即感染倍数很高)中会发生溶源化过程,当营养供应充足时,溶源菌就能生长,从而使噬菌体基因能无限地保留下来。事实上,决定温和噬菌体进入溶源周期的两个条件正是营养缺乏和大剂量的感染。

## 大肠杆菌 $\lambda$ 噬菌体

研究得最清楚的温和噬菌体是大肠杆菌的  $\lambda$  噬菌体,它的溶源周期是本章的重点内容。在此之前,读者首先应阅读一次前一章中所叙述的  $\lambda$  噬菌体的裂解周期。

### 对感染的免疫性

烈性噬菌体形成的噬菌斑是透明的,因为其中心的细菌全被杀死并裂解。可是, $\lambda$  噬

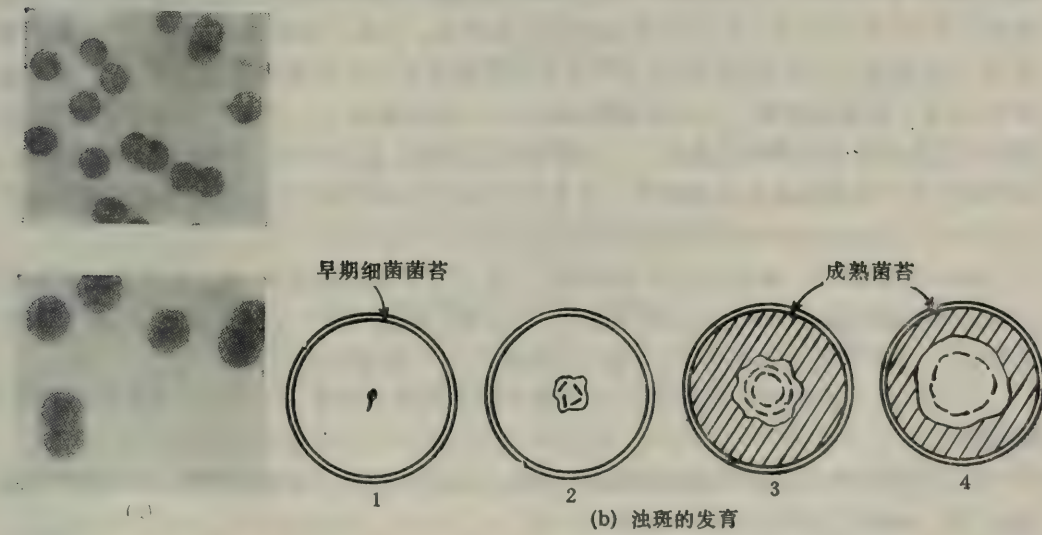


图 16-5 (a)  $\lambda$  噬菌体的清亮透明的噬菌斑 ( $cI^-$ ) 与浊斑 ( $c^+$ )。 (b) 温和噬菌体浊斑发育图示(放大了 5—10 倍): (1) 细菌苔层中的噬菌体; (2) 含有几个溶源菌(以小棒表示)的小的透明斑(通常可见); (3) 透明斑扩大,其中有溶源菌的生长; (4) 透明斑达到最大,溶源菌停止生长(承 A. D. Kaiser 提供)。

菌斑的中心部分是混浊的(其他温和噬菌体也如此)(图 16-3),这是由能抵抗  $\lambda$  感染的溶源菌的生长造成的。通常在铺培养平板时噬菌体和细菌采用  $1:10^5$  的比例,细菌能很快生长,而感染倍数低,因此能保证发生裂解。经过几个裂解周期后,感染倍数增高,一些细菌就溶源化了。此时营养充足不限制噬菌体和细菌的生长,所以绝大多数细菌仍被裂解。当培养基中的营养耗尽后,没有受感染的细胞停止生长,噬菌斑的大小就不再增加。但是,在噬菌斑中一些细菌生长得比较慢,在那里还有营养存在,那些能够抵抗  $\lambda$  继续感染的溶源菌还能继续生长,在噬菌体中形成混浊的中心。

$\lambda$  溶源菌能抵抗  $\lambda$  感染的性质,称为免疫性。下面介绍这个现象的原因, $\lambda$  噬菌体具有一套阻遏物-操纵基因体系(图 16-4)。阻遏基因叫做  $cI$ ;阻遏物蛋白结合在  $oL$  和  $oR$  两个操纵基因上, $oL$ 、 $oR$  分别与  $pL$  和  $pR$  两个启动子相邻,L和R两字母表示按标准方向绘制基因图谱时,两个早期 mRNA 合成的方向:左向或右向(见第十五章图 15-29)。

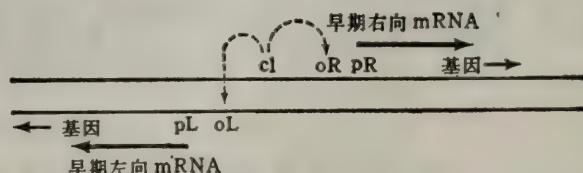


图 16-4  $\lambda$  的阻遏物-操纵基因体系示出了两个早期 mRNA 分子。 $cI$  表示阻遏物基因,P表示启动子,o表示操纵基因,L表示左向,R表示右向。

在溶源菌中, $cI$  阻遏物不断地合成,而且对操纵基因来说稍过量一些。这样, $oL$  及  $oR$  都有结合的阻遏物分子,而使 RNA 聚合酶不能从  $pL$  及  $pR$  启动 mRNA 的合成。因此,在溶源菌中,从两个早期启动子进行的转录是被阻止的。在后面可以看到,这足以使前噬菌体处于“关闭”状态,因此,溶源菌能无限地生长。如果  $\lambda$  噬菌体感染了一个正常的细菌,由于阻遏物还未合成,侵入的  $\lambda$  DNA 的两个操纵基因还未被阻遏物占据,所以能开始转录;如果  $\lambda$  噬菌体感染了一个溶源菌,在 RNA 聚合酶与  $pL$  和  $pR$  位点结合之前,溶源菌中已经存在着过多的阻遏物,这些阻遏物就与侵入的  $\lambda$  DNA 的两个操纵基因结合,这样就阻止了噬菌体进入裂解周期。人们把这种抑制作用称为对于同源免疫超感染抗性(resistance to homoimmune superinfection)。

超感染的 DNA 将出现什么情况呢?一般它可以形成超螺旋,而且在形成超螺旋时不需要噬菌体的基因产物,但不能复制。在这种超螺旋 DNA 存在的情况下,溶源菌不受影响,仍能正常地生长和分裂,于是超感染的 DNA 就逐渐地被稀释了。

已分离到一些免疫体系突变株,其中最重要的两种类型是  $cI^-$  和 vir 烈性(virulent)突变。

1.  $cI^-$  突变株不能合成出有功能的阻遏物,因此它只能进入裂解周期,形成透明的噬菌斑[图 16-3(a)]。

2. vir 突变株的  $oL$  和  $oR$  都发生突变,因为它不能建立阻遏作用,所以 vir 突变株也形成透明的噬菌斑;而且由于它对溶源菌中已有的阻遏物不敏感,所以它能超感染溶源菌并在其中生长。



表 16-1 不同的噬菌体在同源免疫和异源免疫的溶源菌上形成噬菌斑的能力

超感染噬菌体	溶 源 菌		
	B( $\lambda$ )	B(434)	B(21)
$\lambda$	-	+	+
434	+	-	+
21	+	+	-

注: B(P) 表示 P 噬菌体的 B 溶源菌。+代表能成斑,-代表不能成斑。

大肠杆菌除了  $\lambda$  外还有许多种温和噬菌体,与  $\lambda$  有关的两种是  $\lambda$  噬菌体 21 和  $\lambda$  噬菌体 434。这些噬菌体都有自己的免疫系统,即自己的阻遏物和专一于这种阻遏物的操纵基因。因此, $\lambda$  434 的阻遏物不能与  $\lambda$  的操纵基因结合, $\lambda$  的阻遏物也不能与  $\lambda$  434 的操纵基因结合,这样一对噬菌体被称为相互的异源免疫 (heteroimmune)。由于溶源菌中的阻遏物不能与异源超感染噬菌体的操纵基因结合,因此温和型噬菌体能在异源免疫的溶源菌中形成噬菌斑,这些情况总结在表 16-1 中。一些包含着 *ci* 基因、操纵基因、启动子(和其他一些马上就要讨论的成分)的 DNA 的免疫区用基因符号 *imm* 表示,如 *imm $\lambda$* , *imm21*、*imm434* 等。通过对两种异源免疫噬菌体进行重组,再从中挑选出含有一个噬菌体的免疫区基因和另一个噬菌体的其他基因的重组子。这是一种使人很感兴趣的杂交噬菌体,这种噬菌体在实验室中很有用处。一个突出的例子是杂交的  $\lambda$  434 噬菌体。表 16-1 示出在同源、异源免疫溶源菌中不同噬菌体的成斑能力。基因符号为  $\lambda$  *imm434* (有时也表示成  $\lambda$  434hy),如图 16-5 所示。

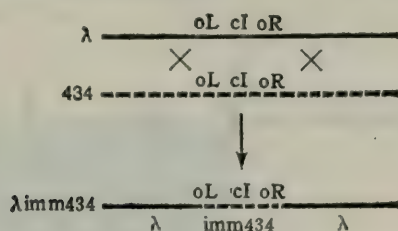


图 16-5  $\lambda$  *imm434* 的形成。

## 前噬菌体的整合

在  $\lambda$  或与其类似的噬菌体的溶源周期中,噬菌体 DNA 分子插入(整合)到细菌染色体 DNA 上,形成前噬菌体。为了形成溶源菌,对噬菌体有两个要求:(1)噬菌体必须建立其阻遏状态;(2)  $\lambda$  必须使其 DNA 分子插入到细菌染色体形成前噬菌体。

$\lambda$  DNA 发生整合时,插到大肠杆菌染色体上的一个优先的位点上,这个位点位于半乳糖操纵子 (*gal*) 和生物素操纵子 (*bio*) 之间,称作  $\lambda$  的附着位点 (attachment site),其遗传基因符号为 *att*。绝大多数温和型噬菌体的整合都发生在一个优先插入的位点上。

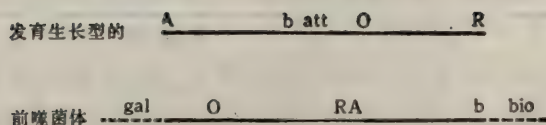


图 16-6 噬菌体头部 DNA (生长发育顺序)和前噬菌体 DNA (前噬菌体顺序)的基因顺序。  
为了标出参照点挑选了一些基因注以字母,其他基因是任意选的。

当然也有几种噬菌体,它们或是有几个插入位点,或者在大肠杆菌染色体的任意位点上都可以插入。对  $\lambda$  前噬菌体的基因顺序的观察,发现插入作用的一个重要特点即基因在前噬菌体中的顺序与在噬菌体中的顺序不一样(图 16-6)。这种情况可以用人们熟知的整合模型来解释,这个模型由 Allen Campbell 提出,就称作 Campbell 模型。在这个恰当的模型(见图 16-7)中, $\lambda$  DNA 首先环化,然后被打开,与宿主 DNA 连接,最后完成前噬菌体的整合。这个过程发生在细菌 DNA 的 att 位点和噬菌体 DNA 上靠近中心的噬菌体的 att 位点,由噬菌体编码的整合酶(integrase, 其基因记作 int)识别噬菌体 DNA 与细菌 DNA 的附着位点,并催化物理性交换,其结果为:噬菌体整合到细菌的 DNA 上,如图所示。环化和整合的结果是感染性的  $\lambda$  DNA 的线性分子的基因顺序改变了。在这个模型提出之后又产生了一个问题,既然前噬菌体两侧有结合位点,那么在整合之后整合酶为什么不很快地将前噬菌体剪切下来,而且,由于一个 DNA 分子内的两个附着位点(宿主 DNA 中)的相互作用应当比不同 DNA 分子(噬菌体 DNA 与细菌 DNA)上的附着位点相互作用得更快一些,所以从动力学上看,应更容易发生剪切反应,为什么没发生呢?当以下情况被发现后,上面这个问题也就清楚了,那就是噬菌体 DNA 和细菌 DNA 中的附着位点是不同的,它们重组到一起,形成前噬菌体的附着位点。这样形成的前噬菌体的附着位点与原来两者都不相同。

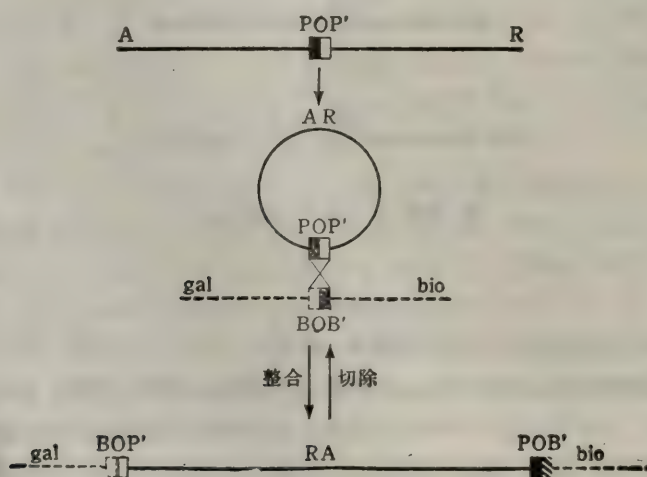
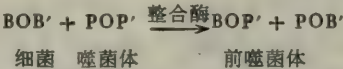


图 16-7  $\lambda$  前噬菌体整合与切除机理的 Campbell 模型。为了与后来的发现一致,噬菌体的附着位点用 POP' 表示。细菌为 BOB',前噬菌体两侧新的结合位点分别是 BOP' 和 POB'。



所有的  $\lambda$  的附着位点都具有三个不同的组分,其中一个存在于所有的附着位点之中,用  $O^*$  表示。噬菌体附着位点写成  $POP'$  ( $P$  代表噬菌体),细菌附着位点写成  $BOB'$  ( $B$  代表细菌)。因此,在整合反应中产生两个新的附着位点(见图 16-7)  $BOP'$  和  $POB'$ 。通常用  $attL$  和  $attR$  分别表示前噬菌体两端的左、右两个附着位点,由于整合酶不能催化  $BOP'$  和  $POB'$  之间的反应,因此下面这个反应在有整合酶时是不可逆的



整合的结果是  $\lambda$  DNA 以线性状态插在细菌染色体 DNA 的  $gal$  与  $bio$  位点之间,作为大肠杆菌 DNA 的一个片段一起复制。第一个插入的证据来自遗传学实验,表明 (1) 在前噬菌体中,基因顺序发生改变; (2) 从  $gal$  到  $POP'$  右边一些基因之间是连锁的; (3) 从  $bio$  到  $POP'$  左边的一些基因也是连锁的; (4) 溶源菌中  $gal$  与  $bio$  之间的距离大于非溶源菌中这两个位点之间的距离。最终的插入证据来自下面的物理实验。将噬菌体与一个分子量为  $105 \times 10^6$  的含有  $BOB'$  位点的环状质粒进行整合形成前噬菌体。质粒仍然保持环状,但其分子量因增加了一个  $\lambda$  分子的分子量 ( $31 \times 10^6$ ) 而达到  $136 \times 10^6$ 。

现在已得到纯化的整合酶,  $BOB'$  和  $POP'$  位点的序列也已知道。这样,在体外就可进行整合反应(图 16-8)。整合酶具有与 DNA 结合的活性,能与  $POP'$  强烈地结合。它同时也是 I 型拓扑酶(见第八章),能将双螺旋中的一条链打断,将断链绕着连续链旋转,

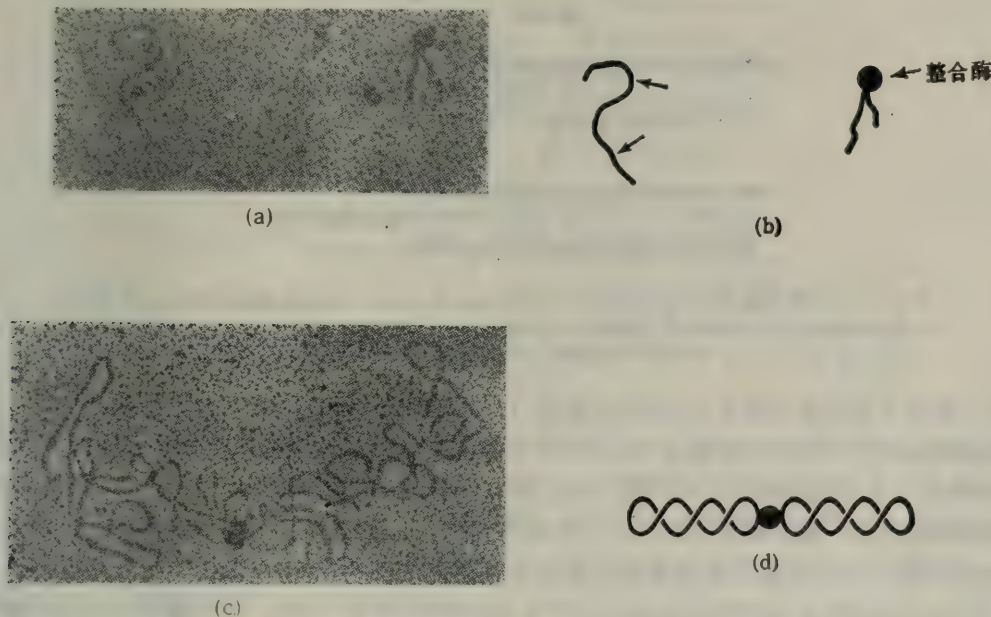


图 16-8 (a) 体外整合酶催化反应的电泳照片。左边是  $POP'$  的  $\lambda$  DNA 片段,右边是另一片段,纯化的整合酶结合在它上面。(b) 是 (a) 图中两个分子的图解,两个箭头表示整合酶结合位点,右边的分子的因上述两个位点(相距 250 碱基对)接近而缩短,说明整合酶把两个位点连在一起了。(c) 两个都带有  $POP'$  位点的超螺旋的  $\lambda$  DNA 分子,由整合酶连接到一起。(d) 是 (c) 图中两个分子的图解(承 Mau Better 提供)。

\* 另一种表示  $BOB'$  的方法是用“●”代替  $O$ ,大肠杆菌整合位点记作  $B \bullet B'$ ,在本书中我们更喜欢用  $EOB'$ 。

然后两个末端连接起来。除了整合酶，整合反应还需一个叫做 IHF 的宿主蛋白 (integrate host factor, 整合酶的宿主因子)，IHF 也结合到整合位点上，但其生化功能还不清楚。IHF 是一个二聚体或四聚体，它的一个亚基是 himA 基因的产物，himA 蛋白也以某种形式参与对溶源途径和裂解途径的选择。序列分析表明，B, B', P 和 P' 的序列差别很大，而通常被称作核心序列 (core sequence) 的 O 序列 (像在 BOB' 中的) 中会有大量 A-T 碱基对 (见图 16-9)。还不详细知道整合反应的机理，但已清楚了几点。首先，如图 16-10 所示，在每个整合区的互补链上相隔 7 个碱基对处打开两个单链缺口；第二，同样如图 16-10 所示，发生交换，同时生成两个由核心序列的链形成的重叠结合接头。最近



图 16-9  $\lambda$ att 区核心序列的碱基顺序。

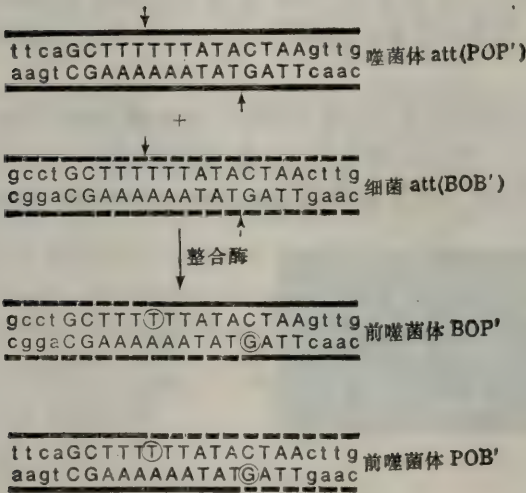


图 16-10 在整合酶催化的交换反应中，DNA 链被酶切断的大约位置(箭头所指)。attL 和 attR 中带圆圈碱基的来源未能确定，也许来自与不同(虚线和实线)的整合区。小写字母表示两侧 B, B', P 或 P' 中的碱基，大写字母表示 O 区中的碱基。

有人提出了整合反应酶反应的模型 (见图 16-11)，认为整合附着位点处是易溶解的。它的富含 A + T 的核心序列解链，与另一个不同的整合位点的互补序列互补，形成四链片段 [步骤 (b)]。在步骤 (c) 中同源区形成具有两个单链缺刻的配对。接着，整合酶的拓扑酶活性使每个双链区旋转 270°。然后，来自两个核心序列的相应序列连接起来。上述“缺刻—旋转—连接”过程又在碱基互补区的另一侧再进行一次 [图步骤 (d)]。然后，带有重叠结合接头的两条链所形成的核心序列又再度分开 [步骤 (e)]。这个模型在大的步骤方面基本上是正确的，但还需再进行更多的实验以证明其细节。

整合酶的合成与 cI 阻遏物的合成相偶联。这样是很有效的，因为在溶源周期中两步都是必需的，而在裂解周期中两者都是不需要的。图 16-12 是对这种偶联进行调节的概要说明。在缺乏正调控因子时，(cII 基因的产物) int 和 cI 基因的启动子都不能与 RNA 聚合酶结合。在感染后不久，cII 蛋白被合成出来 (见第十五章)，当 cII 蛋白浓度足



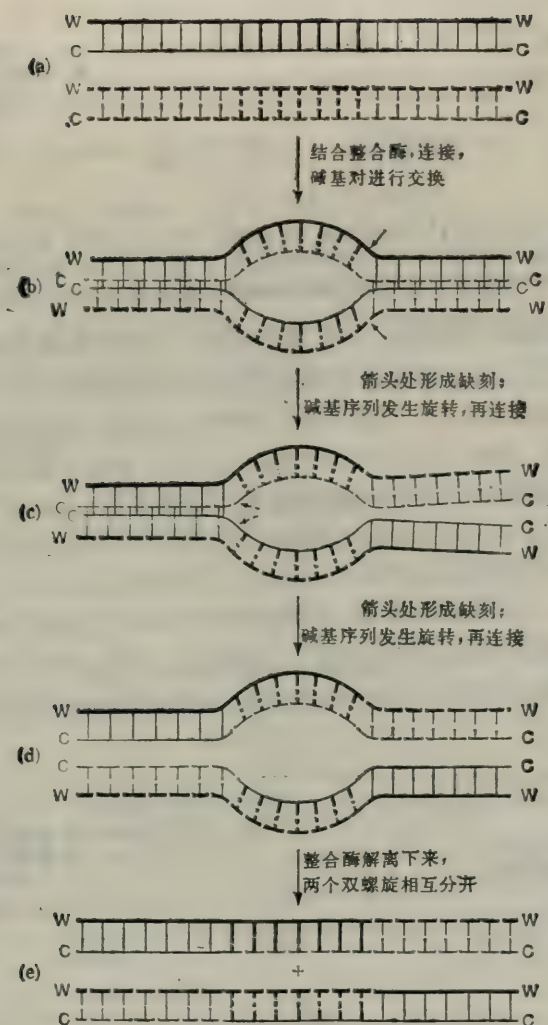


图 16-11 整合反应模型。详见正文。粗线表示打开的碱基对(承 Howard Nash 提供的图,已经过修饰)。

够高时,它结合到 *cI* 基因和 *int* 基因的启动子(分别用 *pre* 及 *pI* 表示)附近的位点上,使 *pre* 和 *pI* 能与 RNA 聚合酶结合(在这一点上, *cII* 蛋白的作用与乳糖操纵子中 *cAMP*-*CAP* 的复合体的作用类似),然后, RNA 聚合酶转录这两个基因,生成这两个基因的产物:阻遏物与整合酶。以后我们会看到,通过调节 *cII* 蛋白的浓度来决定在感染中  $\lambda$  噬菌体是进入溶源周期或是进入裂解周期。

### 诱导及前噬菌体的切除

溶源性细菌一般都很稳定,并且能在不释放出噬菌体的情况下无限地复制。可是,一旦溶源性细菌遭到破坏,便有利于噬菌体,它就脱去阻遏而开始其裂解循环。这种情况确实存在,并且去阻遏作用(前噬菌体诱导作用, *prophage induction*)的信号就是 DNA

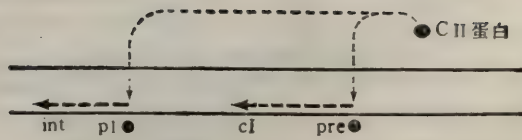


图 16-12 在 pre 与 pl 分别合成出 cl mRNA 和 int mRNA 的过程中 cII 蛋白的作用。带箭头粗虚线表示 mRNA 分子,箭头方向表示 RNA 合成方向,箭头所靠近的 DNA 链是被转录的链。圆点表示 cII 蛋白。细虚线箭头指出 cII 蛋白的结合位点。

损伤。目前还不知道诱导的机理如何,图 16-13 表示了两种目前流行的理论。迄今为止,紫外线照射是所有诱导物中研究得最广泛的,它能导致 DNA 损伤从而引起 SOS 修复系统的活化(见第九章)并且在 DNA 中产生单链区。紫外线照射之后,发生了一系列未知的连续过程,使得 RecA 产物(见第十四章)的蛋白酶活性对 cl 基因编码的阻遏物进行切割而失去活性,从而诱导噬菌体进入裂解循环。RecA 蛋白是 SOS 修复作用和遗传重组作用中的一个比较重要的因子。

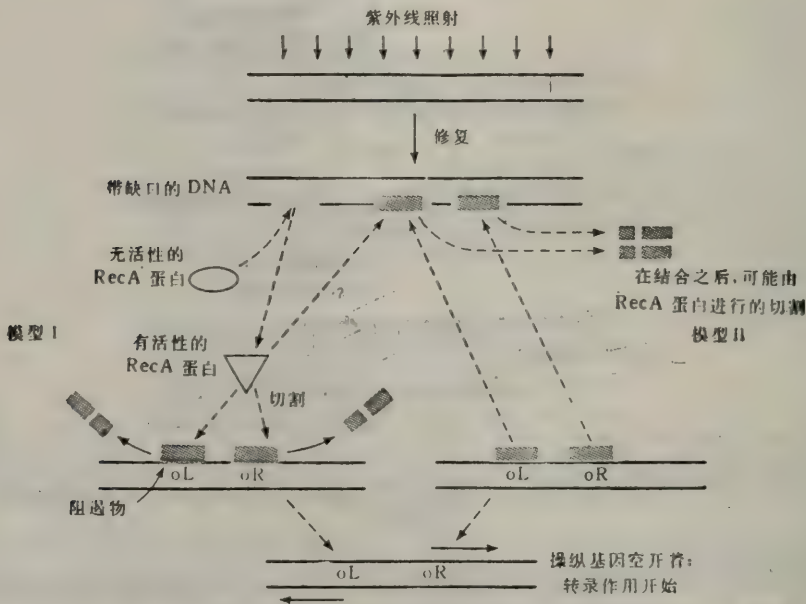
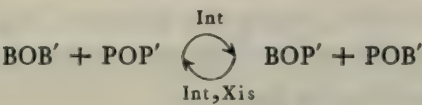


图 16-13 前噬菌体诱导作用的两种途径。灰色表示阻遏物。在第一种模型中,紫外线照射之后的早期步骤是活化 RecA 蛋白,这可能是通过结合到紫外修复作用诱导产生的缺口上而实现的。这个活化的 RecA 蛋白切割阻遏物,使操纵子获得自由。在第二种模型中,由于紧密地结合到有缺口处的 DNA 上,而使阻遏物离开操纵子。

从 pL 开始转录的产物是一种称做切除酶 (excisionase) 的蛋白质(简称为 Xis),它是由 xis 基因编码的。近来已纯化了这个切除酶,但只知道它的少数几种性质。遗传学证据及结合性研究表明,切除酶和整合酶在一起形成一个复合物,于是后者去识别前噬菌体的附着位点 BOP' 和 POB'; 一旦这个复合物与这些位点结合,整合酶即可在核心序列上进行切割,重新形成 BOB' 和 POP' 位点。最近通过电镜研究阐明了 Xis 的物理作



用。电镜研究表明, POB' 环绕 Xis 形成了一个紧密的复合物,据认为,这个复合物中也包括 Int。在 POB' 和 BOP' 相互作用之前就形成这种复合物。于是 Xis-Int 复合物使整合反应逆转,导致前噬菌体被切除(见前面的图 16-7)。所有附着位点之间的反应可以写作



Xis-Int 复合物不能催化右向反应,所以,当切除酶存在过多时,切除就是不可逆反应。

注意,切除的产物是一个完整的大肠杆菌染色体和一个环状的 λ 噬菌体分子,并且在刚受到噬菌体感染的细胞中存在的也是这两种物质。

### 合成 λcI 阻遏物的两种启动子

在溶源周期中很重要的一点,就是比 DNA 复制快的阻遏物分子的合成增加了操纵基因的数量。如果操纵基因的数量对于阻遏物的数量太多了,就会发生转录,导致大量产生噬菌体和细菌死亡。因此,当进入溶源周期而刚被感染后,阻遏系统就会突然地大量合成阻遏物。反之,如果在溶源菌中只含有一个 λDNA 分子(前噬菌体, prophage),维持阻遏状态就只需要较少的 cI 阻遏物。很显然, cI 基因必须在溶源性细菌中才转录,但是转录并不强烈(这是对以前说法的修饰,以前认为在溶源状态时前噬菌体处于“关闭”状态)。

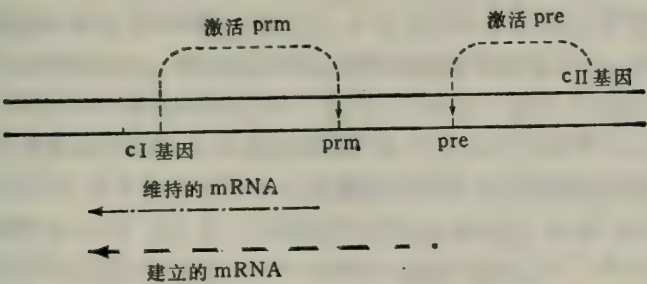


图 16-14 λ 噬菌体的 cI 基因及其邻近的区域,标出了合成 cI mRNA 的两个启动子。两条实线是 DNA 两条链,下面带箭头的虚线代表都从下面的那一条 DNA 链转录出来的两个 mRNA 分子,上面的链向右转录出的 mRNA(图中未标出),翻译出 cII 基因产物,这个蛋白质活化 pre,产生建立溶源状态的 mRNA。cI 蛋白活化 prm,产生维持溶源状态的 mRNA。

在 λ 中,阻遏物的高水平合成和低水平合成是通过两个启动子控制的(见图16-14)。(1) 建立溶源状态的启动子 pre, 它被 cII 蛋白活化,并在被感染的细菌中起作用。(2) 维持溶源状态的启动子 prm,它被自身阻遏物调节并在溶源菌中起作用。因此,当溶源化作用发生之后(也就是当噬菌体已经合成出足够活化 pre 的 cII 蛋白时),开始于启动子 pre 的 mRNA 就会突然地大量合成阻遏物,并且这个阻遏物蛋白质又去活化启动子 prm。当阻遏物的数量增加时, cII 基因的转录受到阻遏,并且不再产生 cII 蛋白。cII 蛋白不稳定,它的浓度很快降到不足以使 pre 维持在有活性的状态,于是 pre mRNA 的合成就终止了。cI 蛋白活化 prm,产生 prm mRNA。prm mRNA 继续不断地合成。这些 mRNA

在以后许多世代都能翻译,这样就使阻遏物维持在一定的浓度水平上,使之足以阻遏前噬菌体。正像我们已经说明的那样,在溶源菌中只需要很少的阻遏物,从效率观点讲从 *prm* 开始的转录应该少一些。但是事实上,两个启动子都处于全开放状态,以同样活跃的程度与 RNA 聚合酶结合。然而由于从 *pre* 转录下来的 mRNA 上有与核糖体结合很强的结合位点,而 *prm* 的转录子却只能与核糖体微弱地结合,所以生成的阻遏物也就较少。因此,翻译效率决定了来自两种 mRNA 的阻遏物的数量不同。

表 16-12 是阻遏物合成过程的总结。

表 16-12 溶源途径中 *cI* 阻遏物合成的大致步骤

1.感染细胞
2.从 <i>pR</i> 向右转录出 mRNA
3.活化 <i>pre</i>
4.从 <i>pre</i> 转录并翻译出阻遏物(高水平合成)
5. <i>cI</i> 阻遏物活化 <i>prm</i>
6.从 <i>prm</i> 转录并翻译出阻遏物
7. <i>cI</i> 蛋白关闭 <i>pR</i>
8. <i>cII</i> 蛋白降解,因此 <i>pre</i> 失活
9.从 <i>prm</i> 继续合成 <i>cI</i> 蛋白(从低水平合成)

我们刚说过需要阻遏物去开动从 *prm* 开始的转录,当阻遏物浓度很高时,从 *prm* 开始的转录不能进行,这是因为阻遏物负调控它本身的转录,这个调节机理使  $\lambda$  阻遏物的浓度维持稳定。这种稳定有两个好处:(1) 由于营养供应波动,自然界中细菌的生长速度也是不断变化的。这种稳定可以使阻遏物的浓度不致于少到自发地发生诱导作用。(2) 如果没有调节,阻遏物的浓度就会高到当需要发生诱导时却不能发生诱导。

也许有人会问,*prm* 为什么会被阻遏物活化? 由于从 *prm* 合成的阻遏物比从 *pre* 合成的少,那么,如果  $\lambda$  只有溶源途径,*prm* mRNA 的合成就是一个必要的过程;但是当噬菌体也有可能进入裂解周期时,就应避免 *prm* mRNA 的组成型合成即非诱导型的合成 (constitutive synthesis),因为这种合成能造成阻遏物的过早成熟,并且这将会阻止启动裂解周期。有了负调控系统,只需控制来自 *pre* 的阻遏物的爆发性合成,就不能从 *prm* 合成阻遏物。

为了理解这些调节循环的机理,需要了解操纵子的结构,这将在下一节中介绍。

操纵基因的结构以及阻遏物的结合和 *Cro* 产物

*oL* 和 *oR* 操纵基因可以进一步分成六个区域: *oL1*, *oL2*, *oL3*, *oR1*,*oR2* 和 *oR3* (见图 16-15)。这些区域的碱基序列很类似,每个区域都可与 *cI* 阻遏物结合,其结合力各不相同,按其强弱顺序如下:

$$\begin{aligned} oL1 &> oL2 > oL3 \\ oR1 &> oR2 > oR3 \end{aligned}$$



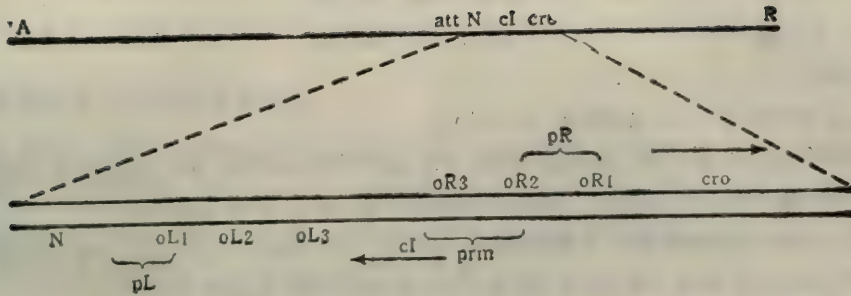


图 16-15  $\lambda$  的操纵基因-启动子区域的详图。箭头方向表示 *cro* 及 *cI* 基因的转录方向。参考图 16-4。

与 *oL1*、*oL2* 和与 *oR1*、*oR2* 的结合是有顺序的,同时又是协同的,与 *oR3* 和 *oL3* 的结合则不是协同的。

当阻遏物的浓度很低时,*oL1*、*oR1* 位点就被阻遏物占据着,由于 *pL*、*pR* 分别与 *oL1*、*oR1* 相邻,因此当阻遏物结合在这些位点上时 RNA 聚合酶与 *pL*、*pR* 结合受到阻碍。当 *oL2*、*oR2* 也结合了阻遏物时,这种堵塞就更加彻底。这时操纵子处于溶源周期的状态。因为 *prim* 启动子位于 *oR2* 与 *oR3* 之间,因此,如果阻遏物浓度很高, *prim* 就被堵塞;这就解释了阻遏物对自身合成的负调控作用。同时,只有当 *oR1* 被阻遏物占据时, *prim* 才能与 RNA 聚合酶结合。*prim* 的活化机理还不清楚,但知道阻遏物与 *oR1*、*oR2* 结合是阻遏物合成的自身正调控方式,可能是阻遏物的结合使 *prim* 中的碱基配对减弱,从而容易被 RNA 聚合酶形成开盒式的启动子复合体,这很像 cAMP-CAP 复合体对乳糖操纵子的正调控(见第十四章)。

因此,如果感染大肠杆菌的  $\lambda$  噬菌体要进入溶源状态,就会发生下面的事件:

1. 开动从 *pL* 和 *pR* 起始的转录。

2. 从 *pR* 开始的转录产物翻译出 *cII* 蛋白。

3. *cII* 蛋白使 RNA 聚合酶能从 *pre* 启动子处开始转录,合成阻遏物。

4. 阻遏物结合到 *oL1* 与 *oR1* 上去,关闭 *cII* 基因的转录,于是 *cII* 蛋白的合成随之停止。*cII* 蛋白是不稳定的,不久,从 *pre* 的转录也就停止了。

5. *oR1* 被阻遏物占据(见第 4 点)使 *prim* 得到活化,这样使阻遏物能从这个启动子合成出来,不过是低水平的合成。

6. *cI* 阻遏物不断地积累着,于是 *oL2*、*oR2* 也被占据了。从 *pL*、*pR* 开始的转录就被完全阻抑了。

7. 如果阻遏物浓度太高,它就和 *oR3* 结合,阻抑从 *prim* 起的转录作用。因此,阻遏物浓度在小范围内发生波动都会影响 *prim* 的活性:使 *prim* 处于“开启”或“关闭”的状态。

我们已经叙述了使得噬菌体不可能进入裂解周期的体系。噬菌体进入裂解周期必须有 *cro* 基因产物参与,这个蛋白质对于裂解周期中的几个步骤是必需的,而且可以防止 *cI* 阻遏物的合成。*Cro* 蛋白也是一种阻遏物,它能与 *oL*、*oR* 操纵基因结合,起作用的方式是基于它对操纵基因中亚位点(subsites)的亲合性有所不同。这种亲合性与 *cI* 对这些亚位点的亲合性正相反:

$$oR3 > oR2 \cong oR1$$

$$oL3 > oL2 > oL1$$

其过程如下:

1. 从 pR 转录 Cro 基因(图 16-15)。
2. Cro 转录之后, cII 基因被转录, cII 基因产物对活化 pre 启动子是必需的。
3. Cro 蛋白合成出来后,立即结合到 oR3 上,阻止 prm 的活化。
4. Cro 蛋白浓度增加,可逐渐地与 oR2、oR1 结合,当 oR2、oR1 全被占据时, RNA 聚合酶就不能与 pR 结合,最终 Cro 蛋白的合成也就停止。

第 3、第 4 步说明, Cro 蛋白在阻抑 cII 蛋白合成的同时,也阻抑了自身的合成。

裂解所需要的大部分基因都在 cro 基因右边,因此,如果 Cro 蛋白是一个高效阻遏物,那么,既不能发生溶源(因为不能产生 cI 阻遏物),也不能发生裂解(缺少从 pR 开始的转录)。然而, Cro 蛋白与 oR 的结合力很弱,当细胞中已有足够量的 Cro 蛋白是去关闭右向的转录以及去阻抑 prm 时,细胞中同样也存在着很多 cII 蛋白,这些 cII 蛋白促进从 pre 开始的转录,并阻抑 Cro 蛋白的合成。如果在早期合成了足够的 cI 阻遏物,那么 Cro 蛋白就不能合成,同时会建立阻遏状态。事实上, Cro 蛋白的合成与 cI 阻遏物的合成之间的竞争决定了噬菌体是进入溶源周期还是进入裂解周期。对这种竞争进行调控的一个严格的调节因子就是 cII 蛋白,在下一节中我们将叙述 cII 蛋白。

### 裂解周期还是溶源周期的决定

决定进入裂解周期还是进入溶源周期的机理还未完全得出\*。但是, cII 蛋白起着核心作用,这是很清楚的。cII 蛋白有三个重要功能: (1) 活化 pre 启动子; (2) 活化 Int 基因的启动子 pI; (3) 延迟晚期的 mRNA 的合成(这个 mRNA 编码噬菌体的头、尾及裂解系统)。有几个证据表明当 MOI 较低时,有活性的 cII 蛋白的含量限制 pre 启动子的活化,其原因还不清楚,但这解释了进入溶源状态需要有很高的 MOI 的原因。有可能活化 pI 和 pre 的 cII 蛋白含有多个亚基,因此,只有当这个蛋白质的总浓度维持在一个很高的水平时,才存在它的活化形式。而这种情况只有在 MOI 很高时才可能出现。

决定进入裂解还是进入溶源周期的两个重要因子是大肠杆菌的 hfl 基因和 himA 基因。很明显, hfl 基因编码(或控制)着可以降解 cII 蛋白的二个蛋白酶中一个。然而它的活性被  $\lambda$  的 cIII 基因蛋白抵销, cIII 基因蛋白则是由起始于 pL 的 mRNA 所编码的。HimA 在溶源化中有两个作用: (1) 如前所述,这个宿主蛋白作为整合酶的一个亚基参与整合反应,它是一个必需的成分; (2) 当对 himA 宿主进行感染时,在这个体系中除了不能合成出整合酶之外,还不能合成出 cI 阻遏物。此外, himA 基因还参与许多细胞内的反应,例如, SOS 修复和遗传重组。HimA 蛋白的量随细胞生长状态的变化而变化。可能是 HimA 蛋白和(或) Hfl 蛋白的浓度随细胞生长变化而发生的变化影响了早期所说的溶源化的要求——细胞生活的环境必需是略微缺乏营养的。至于 HimA 蛋白是如何影响对裂解和溶源的选择,现在还不清楚。但是关于 HimA 蛋白是一个对环境敏感的

\* 这个研究得很活跃的题目还没有完全弄清楚。书中的一些陈述是靠推测得来的。



因子以及它对于整合反应是必需的看法都已经弄得很清楚了。因此，对于整合反应来说，当 HimA 蛋白的浓度过低时，建立阻遏作用，合成整合酶等对噬菌体都是徒劳的。

溶源化作用过程中的切除及其防止

溶源周期中，在 cI 阻遏物合成之前，发生从 pL 起的转录，结果是 int 基因和 xis 基因的转录。如果整合酶 (integrase) 和切除酶 (excisionase) 同时出现，那么已插到染色体中的前噬菌体就会马上被剪切下来，这是溶源化过程中必须避免的。图 16-16 给出了防止这种切除的简单机理。

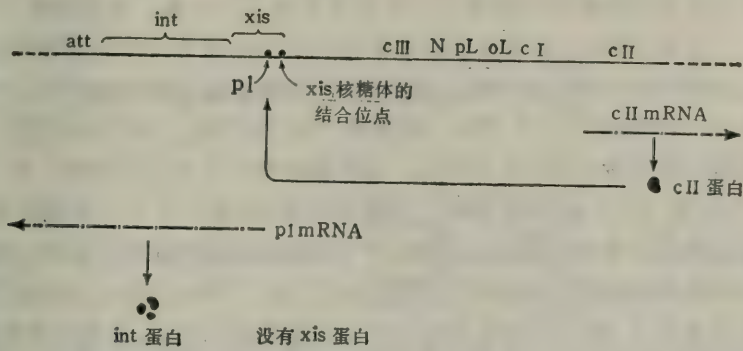


图 16-16 在溶源化过程中防止合成 xis 蛋白。黑线表示 DNA 分子，虚线表示 mRNA 分子，cII 蛋白活化 pI (int 基因的启动子)。pI 位于 Xis 蛋白的核糖体结合位点的下游。

在溶源化途径的一部分过程中，cII 蛋白开动从 pre 启动子起始的 cI 阻遏物的合成，以及从 pI 启动子起始的整合酶的合成。然后 cI 阻遏物作用于 pL，关闭从 pL 开始的 int 及 xis 基因的转录。pI 启动子位于 xis 基因中的核糖体结合位点的下游，这样，mRNA 只能合成整合酶而不是切除酶。

用 λ 筛选致癌物

人们除了对 λ 噬菌体本身感兴趣外，对 λ 噬菌体的认识还为分子生物学研究提供了一种了解基本的分子过程的手段。最近，已证明 λ 噬菌体可用来筛选致癌物。在第九章中描述了一个实验，在这个实验中由于致癌物能诱导沙门氏杆菌中的 His<sup>-</sup> 突变株产生回复突变，从回复突变得知存在着致癌物。人们用含有 λ 噬菌体的溶源化大肠杆菌作了类似的实验(为 Devoret test)，溶源菌以极低的频率自发诱导，而在某些突变株中，这个频率更低。然而，许多致癌剂可以引起 DNA 损伤，并且它们是诱变剂。人们可以很简便地测出这种诱变作用。如果一个溶源菌与 10<sup>8</sup> 个细菌混合，铺到培养基上(像测噬菌斑那样做)，一个没有受到有诱变的溶源菌，在细菌层上形成一个微菌落(最多只有 1 000 个细胞)。偶然，在菌落发育过程的晚期，会有一个细菌被诱变，但是这种诱变通常并不能产生一个噬菌斑，因为培养基已经缺乏营养，限制了细菌生长，所以释放出的噬菌体不能增殖(注意，噬菌体只能在生长着的细菌中增殖)。然而，如果培养基中含有诱变剂(在这种情况下，加的是致癌剂)，当细菌开始在培养基上生长时，就能将初始的那个溶源菌诱变，因

此能够形成噬菌斑。如果在培养基中加入了  $10^5$  个溶源菌,像 Ames 氏测试法那样,通过计量产生噬菌斑细胞的比例就可以知道诱变剂的效率。这个实验为:在培养基中加入肝提取液以活化致癌剂,通过这个试验成功地将这些已在 Ames 氏测试中观察到的致癌剂检测出来。Devoret 氏试验法灵敏度更高,同时,它也只要更低浓度的致癌剂;因此,在 Ames 试验中测不到的弱致癌剂可再用 Devoret 试验法进行检测。这两个试验被用来检测大量的工业生产的化学物质及食品添加剂。

## 其他形式的溶源菌

绝大多数溶源菌按照我们描述的  $\lambda$  的方式溶源化。一般,前噬菌体在细菌染色体上的唯一的一个位点上插入;而能在染色体上好几个 att 位点上插入噬菌体的溶源菌则很少见到。在本章开始时我们曾指出,大肠杆菌 P1 噬菌体显然与此不同,它的前噬菌体不插入到染色体上,感染之后 P1 DNA 被环化,同时被阻抑。在溶源周期中,每一个细胞中含有一至两个超螺旋化的 P1 DNA。在细菌的每一个生活周期中, P1 DNA 只复制一次,并且与细菌 DNA 的复制偶联。当细菌分裂时,每一个子代细胞得到一个 P1 环,但还不清楚这些有序事件的机理,与那些将 DNA 插入到细菌染色体中的噬菌体相比,对于噬菌体如何维持其溶源状态的机理知道的也不多。例如,在每一次细胞分裂时,1000个细胞中有1个丢失了 P1 环,这究竟是由于 P1 偶然地未复制,还是虽然复制了但是没能成功地分配到子细胞中去,尚不清楚。

## 溶源菌的某些性质

当所有敏感的宿主细菌都被噬菌体感染而耗尽之后,溶源化就成为噬菌体繁殖的手段。细菌溶源化的能力对细菌也很有价值,因为它能使细菌在受感染期间存活下来并对同源免疫的超感染作用(homologous superinfection)有免疫性。

溶源化作用对细菌还有其他一些影响,即溶源菌获得原来细菌所没有的一些性质,这称做溶源化转变(lysogenic conversion)。其中一个例子就是  $\lambda$  噬菌体为大肠杆菌提供对 T4 噬菌体的  $\gamma$ II 基因的突变的抗性,一个也是从 prm 启动子开始转录的  $\lambda$  基因——rex 阻断了这些突变株感染的早期阶段。再如在白喉杆菌(*Corynebacterium diphtheriae*)中,只有当细菌成为  $\beta$  噬菌体的溶源菌时,才能引起白喉。噬菌体的插入位点可能在细菌基因内,所以前噬菌体的形成活化了这个基因。具有这种性质的例子还有 Mu 噬菌体,它可在任何基因中形成前噬菌体,所以它是强突变剂;在下一节将要讨论 Mu 噬菌体。

## Mu——一种不寻常的噬菌体

大肠杆菌的 Mu 噬菌体是一种温和噬菌体,它既能溶源化,也能裂解细菌。Mu 这个字来自突变者(mutator),因为这种噬菌体的溶源菌经常地发生突变,所以采用 Mu 来命名。它容易发生突变的原因,是在 Mu 的溶源途径中, Mu 可随机地插入到大肠杆菌染



染色体上,而这些插入位点经常地处于大肠杆菌的一些基因中或在调控序列上,当 Mu 前噬菌体的基因与调控序列分开时,调控序列也就失活了。Mu 的另一个特性在于,它插入到宿主的染色体上的过程,是它的裂解生活周期中一个必要的步骤。鸟类中的逆转录病毒也有这个特性(第二十一章),这个 RNA 病毒引起鸟类的肿瘤。

### Mu DNA

Mu DNA 具有 38 000 个碱基对,是线性的 DNA。它既不是末端冗余的,也不是环状的。然而在许多 Mu 噬菌体中,每个 DNA 分子都互不相同。将 Mu DNA 变性再复性后就会证明这点。来自不同双螺旋的单链分子经复性后,用电镜观察得出图 16-17 所

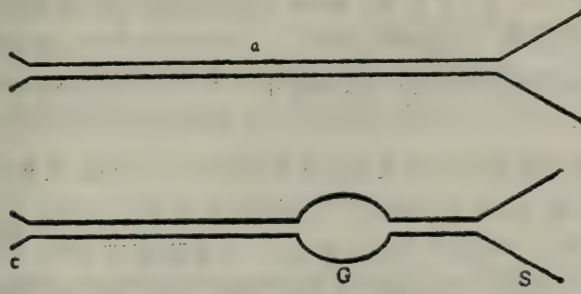


图 16-17 Mu DNA 变性再复性后看到的主要图形。 $\alpha$ 、C、G 及 S 等符号表示杂交双链中可以区分的区域。

示的图形。所有单链都含有不互补的序列,进一步的实验说明了以下特点:

1. 末端的非互补序列是细菌的 DNA 片段,因此,一个 DNA 分子的末端和下一个 DNA 分子的末端不同。
2. 如果 Mu DNA 的一个片段(如在  $\alpha$  区内)的一段长度增加了(由于有外源 DNA 的插入),左端的非同源区(c)的长度保持野生的该区的长度,那么右端的(S)区就会变短。
3. 复性后可能产生一个非互补的称作 G 泡(G loop)的区域, G 泡中两条链是一样的,都是噬菌体 DNA,只是互相之间方向相反,会出现图 16-18(a)所示的那种 G 泡结构。如果 DNA 浓度越低,来自不同链的单链退火就越不能形成杂交双链。不允许单链

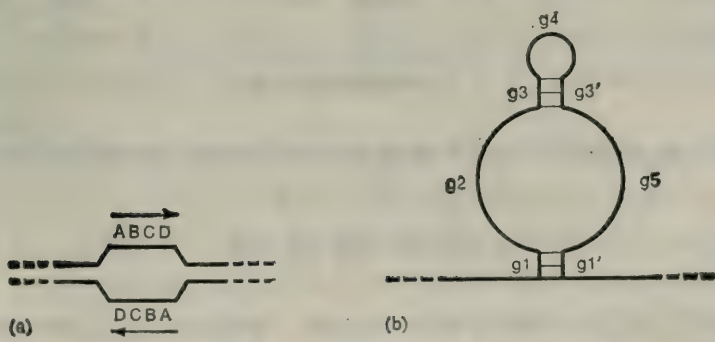


图 16-18 (a) Mu DNA 的 G 泡结构(图 16-17), ABCD 表示 G 序列的倒转。(b) Mu DNA 一条链的退火结果,表示了 G 区中的倒转重复。图中只示出一部分链。

DNA 在 DNA 低浓度下发生自我退火,那么就观察到图 16-18(b) 所示的那种情况,  $g_1$  与  $g'_1$  是自我互补的,在噬菌体生活周期中由于遗传重组 G 片段可能倒转,正常与倒转的构型称作“抛与拍”(“flip and flop”),含有 flip 的链与一个互补的含有 flop 的链退火,产生一个 G 泡。含有 flop 片段的噬菌体的内部 DNA 与其附近的噬菌体的联系发生了改变,因此有 flip-DNA 的噬菌体和有 flop-DNA 的噬菌体产生出不同的尾纤维:它的 DNA 与相邻的噬菌体的启动子的关系不同了,这样就扩大了噬菌体对细菌的可吸附范围。

## Mu DNA 的复制与成熟

对上一节中的第 1,2 点,可用 Mu DNA 的特殊的复制方式来加以解释。

在正常感染中的 DNA 复制期间,子代 Mu DNA 分子插入到含有大量插入位点的染色体上。这个还未弄清楚的现象可以解释第 1 点中所讲的 Mu DNA 与大量的细菌 DNA 相连的现象。

噬菌体 DNA 可能是按下列方式进行包装的(图 16-19)。成熟体系识别 Mu DNA 的 C 末端,并在噬菌体 DNA 的左端 100 碱基对处将宿主 DNA 切断,以这点开始包装,并将头部装满。第二个切割位点在 S 末端,其位置取决于 DNA 装满头部的方式(T4 也用此方式),但一个装满的头部中装着超过噬菌体 DNA 长度的 DNA,这就解释了第 2 点。因此,其成熟体系没有什么特别异常的,只是 DNA 复制的方式很独特(即子代 DNA 插入到染色体中),造成噬菌体含有不常见的 DNA 结构。

在第十九章讨论转座子时将进一步详述 Mu 噬菌体。

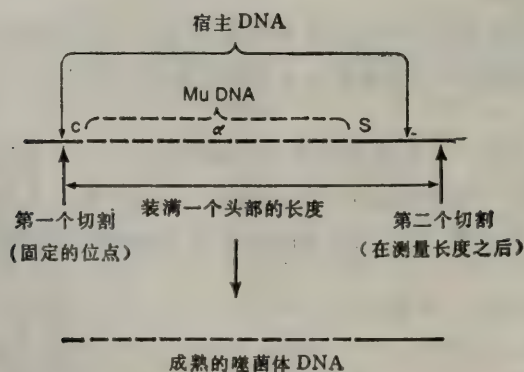


图 16-19 插入的 Mu DNA 的图示(虚线),表明切割过程如何在末端产生出宿主的 DNA 序列。

## 转导噬菌体

到目前为止,在已讨论的噬菌体系统中,DNA 包装进头部的方式有两种。第一种是 T4 噬菌体所采用的装满的方式(见第十五章),噬菌体头部装进定量的 DNA;第二种是确定尾端方式:在串联着的 DNA 的固定位点上造成两个切口。 $\lambda$  噬菌体就是采用后一



种方式（亦见第十五章）。在装满方式中，我们丝毫没讲如果提供自由末端去启动包装，就能防止将宿主 DNA 包进去。由于噬菌体编码的核酸酶能将宿主 DNA 快速地降解，所以刚才那种情形是很少见到的。开始包装时，留下很少量的宿主 DNA，或者根本没有留下。对于  $\lambda$  噬菌体，由于宿主不可能含有适当间隔开的相当于 cos 位点的碱基序列，所以  $\lambda$  噬菌体极少将宿主 DNA 包装进去。但我们还会看到有时  $\lambda$  DNA 带着一些宿主 DNA 而被包装的事件。

现已知道的几种能包装宿主 DNA 的噬菌体，人们把它们称作转导噬菌体 (transducing phage)。含有宿主 DNA 的噬菌体颗粒被称作转导颗粒。这样的噬菌体也分为两类：通用转导噬菌体 (generalized transducing phage)，它能产生出只含有细菌宿主 DNA 的颗粒；第二种是特化的转导噬菌体 (specialized transducing phages)，它偶然能产生出同时含有噬菌体 DNA 与细菌 DNA 的颗粒。两种转导颗粒都能将它们的 DNA 注入到宿主细菌中，并且因此而将宿主的 DNA 由一个细菌转移到另一个细菌中，这个过程就称作转导。特化的转导颗粒的形成是溶源化的结果，将两者放在一起进行讨论是很方便的。

### 通用转导颗粒的性质

P1 感染大肠杆菌所产生的颗粒就是通用转导颗粒的一个很好的典型例子。P1 噬菌体既有溶源周期，又有裂解周期。有裂解周期与形成通用转导颗粒相关。P1 的裂解周期与第十五章中所讨论过的裂解噬菌体没有很大不同。与这个讨论有关的性质是：(1) 每次释放的数目大约是 100 个子代 P1；(2) 噬菌体 DNA 的分子量是  $66 \times 10^6$ ；(3) P1 编码一个核酸酶，它能降解宿主 DNA。与 T4 的核酸酶相比较，P1 的核酸酶的作用非常慢。因此，当包装开始时，宿主 DNA 成为大部分分子量为  $10^7-10^8$  的一些片段，在被感染的细菌中大约有百分之一的宿主 DNA 片段被包装到噬菌体的头部中。P1 的包装体系不是很严格的，因此，所有颗粒中的 DNA 分子的大小不完全相同。而且，头部的装配过程能装配出两种不同大小的头。其结果受噬菌体感染的细菌裂解，产生的大量噬菌体颗粒有一些通用转导颗粒，大部分通用转导颗粒中含有一个分子量为  $20-25 \times 10^6$  的 DNA 片段。由于宿主 DNA 被随机地打成片段，而且这些转导颗粒含有宿主 DNA 的任何一个片段。因此，在一个数量足够大的 P1 噬菌体群体中，至少会有一个具有全部宿主基因的颗粒。对于任何特定基因来说，大约每  $10^6$  个存活的噬菌体中就有一个转导颗粒，因为通用转导颗粒不含有 P1 DNA，所以它不能产生 P1 噬菌体的子代。但是，在有启动子存在的情况下，当通用噬菌体中的细菌 DNA 被注入到宿主细胞中时，基因就能被表达。

我们现在讨论转导颗粒侵染细菌时发生的事件。设想一个来自受感染的野生型大肠杆菌的转导颗粒中含有合成亮氨酸的基因，这样的颗粒称作  $leu^+$  颗粒。当这个颗粒吸附到基因型为  $leu^-$  的细菌上时，就将 DNA 注入细菌，有时细菌能生存下来；一种可能是注入的一段 DNA 不含有噬菌体的基因，而且这段线性 DNA 逐渐地被细菌的核酸外切酶降解；另一种可能是由于基因重组，噬菌体的  $leu^+$  片段整合到宿主 DNA 上，取代了宿主的等位基因  $leu^-$ ，这样，宿主的基因型就会从  $leu^-$  变为  $leu^+$ 。转导颗粒就是这样被检测出并被发现的。也就是说，用噬菌体颗粒感染  $X^+$  细菌培养物，引起细菌裂解，

再用释放出来的颗粒去感染  $X^-$  细菌;然后将这份受感染的  $X^-$  细菌培养物接种在无  $X$  的培养基上。如果培养基上生长出菌落,那么这种细菌一定是  $X^+$  的,一定发生了转导过程,图 16-20 说明了这整个过程。

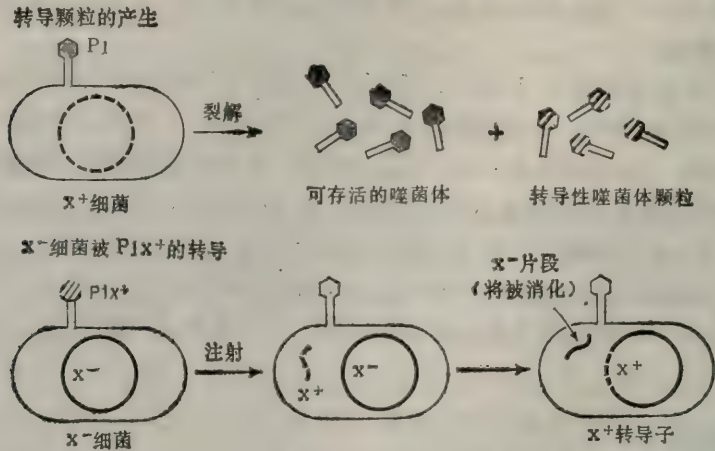


图 16-20 通过 P1 转导使  $X^-$  细菌向  $X^+$  的基因型转变,在上半部分图中,转导性颗粒的发生率为  $1/10^6$ 。

另一种研究得较彻底的是产生通用转导颗粒的是可以感染鼠伤寒沙门氏菌(*Salmonella typhimurium*) 的 P22 噬菌体。对此噬菌体进行的研究为人们提供了转导颗粒只含有细菌 DNA 的证据。这个实验运用了 Matthew Meselson 和 Franklin Stahl 的密度标记技术,图 16-21 示出了其步骤。将细菌在  $^{15}\text{N}$  培养基中生长很多代,然后用  $^{14}\text{N}$  标记的 P22 去感染,并将细菌培养在  $^{14}\text{N}$  的培养基中,再将噬菌体子代进行  $\text{CsCl}$  密度梯度平衡离心,颗粒按密度大小分开,将离心管底部打一个孔,分部收集滴出的液体。这样就将离心后的产物分开了。检测每个收集到的部分中的存活着的噬菌体和转导颗粒。存活的噬菌体具有  $\text{N}^{14}$ -标记 DNA 的颗粒密度(向较高的密度略有一些漂移,因为有一些掺入的  $\text{N}^{14}$  标记 DNA 是从降解了的细菌 DNA 中得到的,所以 DNA 密度稍大一些),但所有转导颗粒都位于含有  $\text{N}^{15}$  标记的颗粒密度的区域,说明它们很少带有噬菌体 DNA。

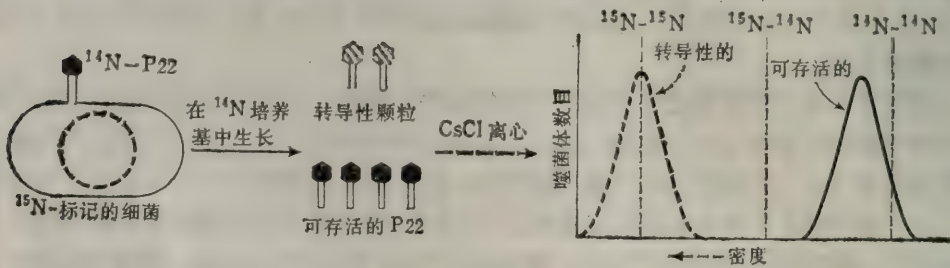


图 16-21 证明 P22 转导颗粒只含有细菌 DNA 的实验。噬菌体 DNA 和细菌 DNA 分别用虚线与实线表示。



在 $\lambda$ 噬菌体感染大肠杆菌后的裂解产物中,也发现过通用转导颗粒,虽然形成的这些颗粒是没有尾巴的。能够形成通用转导颗粒,是因为噬菌体中有 Ter 体系,它负责切割串联的 $\lambda$ DNA 并使之形成粘性末端,它在低频率下偶尔有可能识别细菌 DNA 的碱基顺序,将细菌 DNA 进行部分降解,切成片段而包装起来。所有带尾巴的噬菌体的颗粒都是 DNA 先充满头部,然后尾巴再附在已装满 DNA 的头部上。对于 $\lambda$ 噬菌体来说,如果没有从头部伸出的 $\lambda$ DNA 的左端粘性末端的几个核苷酸,那么其尾巴就不能粘附到它的头部。推测可能是要与头部结合,尾巴必须与特定碱基序列相互作用。然而,在转导粒子中的细菌 DNA 没有正确的碱基序列,所以尾巴不能与头部结合,形成无尾的颗粒。在体外实验条件下,如果采用合适的条件可将 $\lambda$ 尾巴与充满了 DNA 的头部结合,这样可制备出含有细菌 DNA 的通用转导颗粒。

### 通用转导颗粒的用途

由于通用转导颗粒能用来形成新的细菌菌株和绘制基因,它是很有价值的实验工具。最简单的用途就是将正的等位基因引入到缺乏这个基因的(即有负的等位基因的)细菌菌株中。例如,将表现型为  $Leu^- Lac^-$  的菌株转变成  $Leu^- Lac^+$  表现型。首先在任一  $lac^+$  株上培养 P1,裂解液中含有少量的  $lac^+$  P1 颗粒,用它来转导  $Leu^- Lac^-$  菌株可使其转变成  $Leu^- Lac^+$ 。为了进行筛选,将被感染的细菌培养物放在以乳糖作为唯一碳源的培养基上,这样可以将  $Lac^+$  的转导子菌落挑选出来,这就是具有  $Leu^- Lac^+$  表现型的菌株。

转导也可以将负的等位基因引入到带有正的等位基因的细菌中。由于转导后带上了负的等位基因的细菌,不能再在选择性培养基上生长并形成菌落,于是无法直接挑选菌落,所以必须采用间接的筛选方法。所用的方法是在第一章描述的做基因图谱时的连锁过程,下一段将介绍如何发现及检测出连锁以及怎样将它用到转导体系中。

用 P1 去感染  $leu^+gal^+bio^+$  细菌,制备裂解液。此裂解液中的颗粒可转导任何含有这些等位基因的细菌。也就是说,  $leu^-$  细菌可被转变成  $leu^+$ ,或  $bio^-$  可转变成  $bio^+$ 。如果转导  $leu^-bio^+$  菌株,就能产生  $leu^+bio^+$  和  $leu^-bio^+$  两种子代。然而不会产生出  $leu^+bio^+$  细菌。因为一个细菌极不可能同时被  $leu^+$  转导颗粒和  $bio^+$  转导颗粒两种一起感染(这是由于通用转导作用的频率很低所致)。对于  $gal^-bio^-$  细菌来说,情况就不同了。也就是说,不仅可以产生  $gal^+bio^-$  和  $gal^-bio^+$  两种转导子,而且也会产生出  $gal^+bio^+$  细菌,并且是以几乎相同于转导一个单个的等位基因的频率产生的。这是因为  $gal$  和  $bio$  基因之间只间隔着一段分子量为  $15 \times 10^6$  的 DNA。因为转导的 DNA 的分子量可达  $20-25 \times 10^6$ , 这样就有可能装进一段同时含有  $gal$  和  $bio$  基因的 DNA,当出现这种现象时,人们就说  $gal$  与  $bio$  基因是连锁(linkage)的(见图 16-22)。当然,并不是所有的  $gal^+$  转导颗粒都是  $bio^+$  的。因为包装的 DNA 从细菌上切下时,其切点有时位于两个基因之间。很明显,连锁可能性的大小决定于基因之间的距离,距离越近,连锁的机率就越大。对于  $gal-bio$  两基因来说,约有一半的转导子的  $gal^+$  是  $gal^+bio^+$ ,另一半是  $gal^+bio^-$ 。 $bio$  也很相似,  $bio^+$  的转导子中也有  $bio^+gal^+$  与  $bio^+gal^-$  两种,

现在我们来看看连锁过程是如何将一个负的等位基因引入到一个细菌中的。设想有一个很难分离得到的  $x^-$  突变株,它与  $b$  基因连锁而不与  $a$  基因连锁,我们想要得到是基

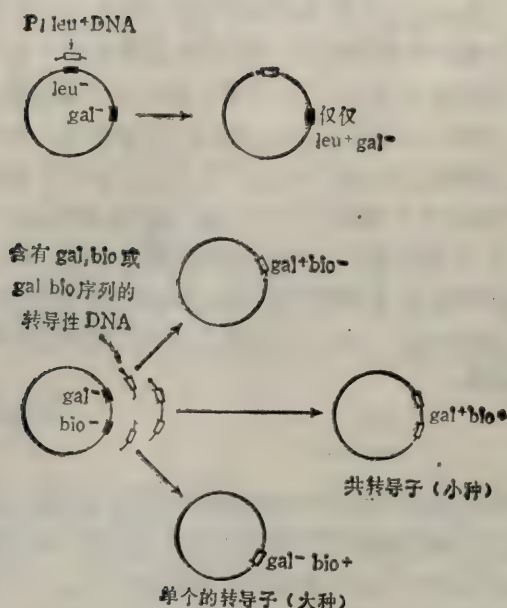


图 16-22 gal 与 bio 基因由于共转导而连锁的图示。

因型为  $a^-b^+$  的菌株。如果  $b^-$  突变株很易分离得到,那么我们就先分离出  $a^-b^-$ 。然后用 P1 去感染  $b^+x^-$  细菌,再将 P1 噬菌体  $b^+x^-$  细菌的培养物(其中有裂解 P1)去转导  $a^-b^-$  细菌,于是细菌转变成  $a^-b^+$ 。由于 b 与 x 基因是连锁的,所以所得到的  $a^-b^+$  的菌株中有一些也是  $a^-x^-$  的。

### 特化转导颗粒的性质

到目前为止,在所讨论的通用转导体系中,裂解周期由于包装了宿主 DNA 片段,就形成了转导颗粒。由于宿主 DNA 切成的片段的过程是随机的,如果裂解产生的噬菌体数目很大,在不均一的通用转导颗粒的群体中就会含有宿主中的各种序列。

已经整合的前噬菌体在其切除作用期间也可能转变,产生出各种不同类型的转导颗粒。这些转导颗粒与通用转导颗粒有三方面不同:(1) 在一个连续分子中,含有相连的宿主 DNA 与噬菌体 DNA;(2) 在这种颗粒中只含有宿主 DNA 的一个或最多有两个区域,特别是那些处于前噬菌体两翼的区域;(3) 可用一个转导颗粒作为产生均一性群体或相同的转导颗粒群的模板。这些颗粒称为特化的转导颗粒。

### 从 $\lambda$ 溶源菌形成特化转导颗粒

图 16-23 示出了形成特化转导颗粒的机理。说明了  $\lambda$  噬菌体的乳糖与生物素转导子(称为  $\lambda gal$  与  $\lambda bio$ )的形成过程。

$\lambda$  前噬菌体被诱导后,紧接着发生一系列变化。前噬菌体 DNA 从宿主 DNA 上被精确地剪切下来,这是由 int 基因及 xis 基因的产物共同地作用于左、右两个前噬菌体的整合位点而完成的。但是,错误的切除也会发生,尽管其频率很低( $10^5$ — $10^7$  个细胞中



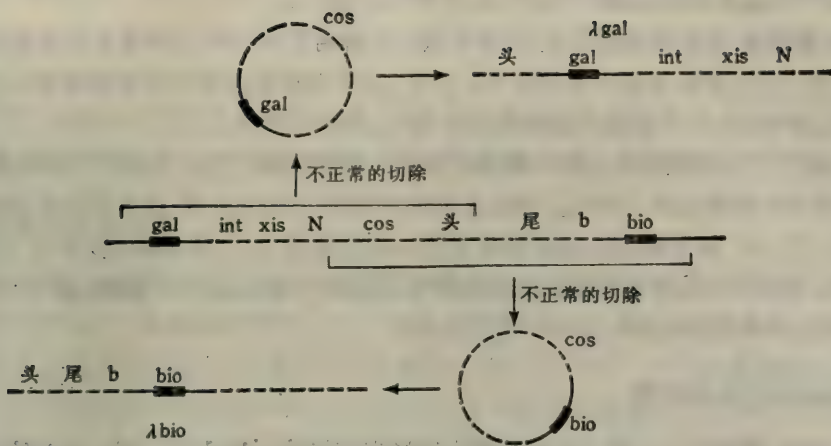


图 16-23 错误的切除导致产生  $\lambda gal$  与  $\lambda bio$  噬菌体。

才会有一个细胞发生错误的切除);有时发生两个错误的切除,一个切点位于前噬菌体中,另一个在细菌 DNA 中。这两个切点不能总是产生适于装入头部的合适长度的 DNA,有时太长,有时太短。只有当长度是正常  $\lambda$  噬菌体 DNA 的 79—106% 时,才能进行包装。由于前噬菌体位于大肠杆菌的 *gal* 和 *bio* 基因之间;同时,宿主 DNA 中的切点既可以在前噬菌体的左边又可以在右边,因此,能产生出带有 *bio* 基因(切在右边)或带有 *gal* 基因(切在左边)的转导颗粒。但是  $\lambda gal$ 、 $\lambda bio$  转导颗粒的形成引起  $\lambda$  基因丢失。 $\lambda gal$  颗粒丢失了位于前噬菌体右端的尾基因,  $\lambda bio$  颗粒丢失了位于前噬菌体左端的 *int*, *xis*……等基因。丢失的噬菌体基因数目取决于生成转导颗粒时所切的位置以及粒子中包含的细菌 DNA 的量。由于丢失的噬菌体基因总是位于前噬菌体的两端,像图中所示的那样,又由于前噬菌体与噬菌体中的基因顺序是可以交换的,所以对噬菌体来说丢失的基因来自它的中央部分。通常情况下,“缺失”(deletion)是指在 DNA 分子中某些野生型的序列不存在了;然而,在特化的转导颗粒中,细菌基因取代了失去的噬菌体 DNA,因此把它们称作为“缺失-取代颗粒”(deletion-substitution particles)。

### $\lambda gal$ 和 $\lambda bio$ 颗粒的性质

由于特化的转导颗粒缺少一些噬菌体的基因,因此,它们可能是不能传代的。 $\lambda gal$  的确由于缺乏合成噬菌体尾部的必需基因,它不能传代,而那些更大些的缺失-取代颗粒,即头部基因发生的大量的缺失,也不能生存下去,它们不能产生子代。我们以字母 *d* 表示这种缺失现象,因此 *gal* 转导颗粒写作  $\lambda dgal$  (有时也写成  $\lambda dg$ )。这种  $\lambda dgal$  颗粒如果含有粘性末端以及所有的用于 DNA 复制与转录的信息,那么它就能够完成一个正常的生活周期,包括细菌裂解。事实上,如果这种转导颗粒没有缺失头部基因,由滚动环产生的串联状的 DNA 分枝能够被 *Ter* 体系切割下来而进入头部进行包装;但是,由于尾巴不能与装满了 DNA 的头部相接,在裂解液中也就不能找到可以生存下去的噬菌体颗粒;在细菌层上也就不能形成噬菌斑。注意,这种  $\lambda dgal$  转导颗粒仍能复制,这种缺失和它

的复制能力之间没有矛盾。只有在也缺少尾巴基因的情况下,  $\lambda$ gal 才不能复制, 但是这样的颗粒只由具有全套基因的正常前噬菌体衍生而来。

$\lambda$ bio 颗粒的情况就不同了, 它通常只缺少非必需基因 *int*, *xis* 等。这些基因在溶原化中起作用, 但在裂解周期中不起作用, 因此, 这些颗粒能复制并形成噬菌斑。为了表示这种颗粒, 就在  $\lambda$ bio 中加上代表成斑的字母 *p*, 写作  $\lambda$ pbio。

*bio* 基因可以将包含着  $\lambda$  的必需基因 *N* 的区域都替换下去, 替换后的颗粒是有缺陷的, 因此将这种颗粒记作  $\lambda$ dbio。通过基因操作可以去掉 *gal* 操纵子与前噬菌体之间的宿主基因, 把 *gal* 基因移到离前噬菌体更近一些的位置上。对这样的溶源性  $\lambda$ gal 颗粒进行分离, 得到了只缺少前噬菌体最右端的非必需的 *b* 基因的  $\lambda$ gal 颗粒(图 16-23), 这种  $\lambda$ gal 颗粒能形成噬菌斑, 用  $\lambda$ pgal 记号表示。

### 非溶源菌的特化转导作用

特化的转导颗粒可以通过几种机理去转导突变细菌。 $\lambda$ gal 与  $\lambda$ bio 颗粒的转导机理是一样的, 所以我们仅用  $\lambda$ gal 作为例子来说明。先对  $\lambda$  噬菌体的  $\text{Gal}^+$  溶源菌进行诱导, 再用诱导产生的裂解液去感染  $\text{Gal}^-$  细胞。这种裂解液中绝大多数是正常的噬菌体, 少部分是  $\lambda$ dgal $^+$  噬菌体。采用可以建立免疫阻遏作用 (immunity repression) 的实验条件, 这样噬菌体就不能通过裂解周期而将宿主细胞杀死。如果将被感染的细菌涂布在一种特殊的琼脂培养基上,  $\text{Gal}^+$  菌落是紫色的, 能从  $\text{Gal}^-$  菌落中区别开来(在该种培养基上  $\text{Gal}^+$  是紫色菌落,  $\text{Gal}^-$  是白色菌落)。紫色菌落出现的频率很低(约占被感染细胞的 0.001%)。这些紫色菌落又可分为下述两种类型(见图 16-24)。

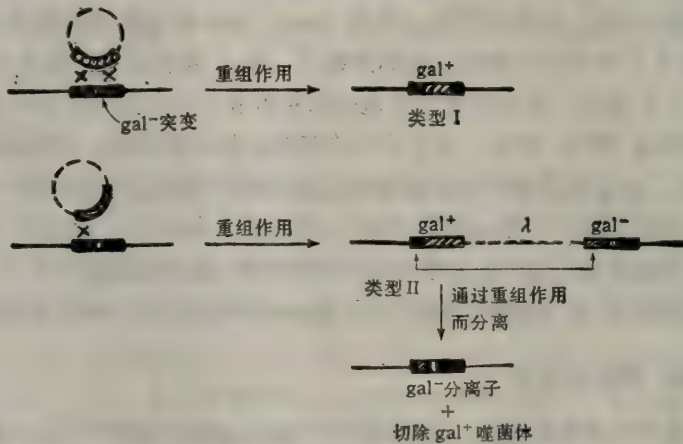


图 16-24 类型 I  $\lambda$ gal 转导子与类型 II  $\lambda$ gal 转导子的产生, 以及从 II 型细胞中分离出  $\text{gal}^-$  细胞。

I 型由非溶源性细胞组成, 全都是  $\text{Gal}^-$ 。如果将菌稀释, 使单个细菌都生成菌落, 形成的菌落都是紫色的。这些菌落是含有稳定的  $\text{gal}^+$  基因型的细菌, 它们是由两个 *gal* 基因进行遗传学双交换产生的, 如图中所示。

II 型细胞含有前噬菌体, 都是  $\text{Gal}^+$ 。如果将 II 型菌稀释进行单菌落涂布培养, 约有 1% 的菌落是白色的(因此是  $\text{Gal}^-$  的), 而且不含有前噬菌体。II 型细胞是由于 *gal* 基因的一个单交换产生的, 因此, 现在这种细菌含有两个拷贝的 *gal* 基因, 即一个  $\text{gal}^+$  和一个  $\text{gal}^-$  基因。这些转导子称作杂合子 (heterogenotes), 而那些只含一个  $\text{gal}^+$  基因的 I 型细胞则都



是单倍体。

当一个细胞中含有两个拷贝的 *gal* 基因时，能按照中等效率发生分子内的遗传重组。这可以说明图中所示的情况——在 II 型 *gal*<sup>+</sup> 转导子连续生长过程中，会产生非溶源性的 *Gal*<sup>-</sup> 细菌。

由于前噬菌体缺乏噬菌体尾部基因，所以 II 型细菌不能被诱导产生子代  $\lambda$ dgal 颗粒。但是，许多噬菌体的专一性蛋白质的基因转录及蛋白质合成都确实在进行着。

### 溶源菌的特化转导作用

$\lambda$  溶源菌可以被转导形成杂合子。这种形成过程以两种方式高频率地发生着。第一种机理与产生 II 型细胞的机理一样，即单个的遗传交换，这种交换或发生在 *gal* 操纵子上或发生在前噬菌体上。注意，如果前噬菌体具有同  $\lambda$ dgal 一样的免疫性阻遏物，就不需再给予引起溶源化反应的条件，因为不会发生噬菌体大量增殖，这些转导子都是杂合子——对于 *gal* 操纵子和前噬菌体两者来说，都是如此。他们与前面所讲的 II 型转导子相比，较不稳定。因为引起产生 *gal* 分离子(segregant)的分子内遗传交换的概率较大，在 *gal* 操纵子中与前噬菌体 DNA 片段中也有这种交换，但是概率比上面那种小一些。

当前噬菌体和  $\lambda$ dgal 是异源免疫(heteroimmune)，具有不同的阻遏物时，发生第二种类型的转导。在这种情况下， $\lambda$ dgal 中的 *int* 基因表达，在前噬菌体与感染性的噬菌体的整合位点之间发生一个单个的遗传交换(见图 16-25)。如图所示，这同样产生不稳定的杂合子(注意， $\lambda$ pbio 中不发生 *int* 介导的所有过程，因为 *bio* 替换了 *int* 基因)。

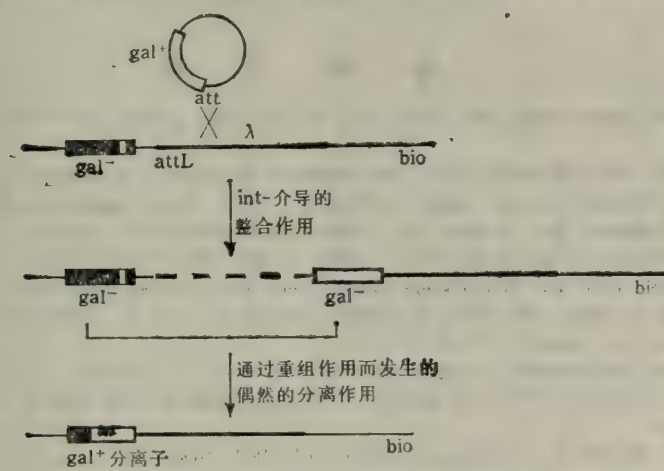


图 16-25 通过前噬菌体整合作用而进行的转导作用。由于遗传重组，偶然发生 *gal* 基因的分离。如果形成某些整合酶，这种分离也会发生在 *att* 位点上。感染性的噬菌体与前噬菌体对于插入作用的发生必须是异源免疫的。

### 高频转导裂解物

上述所讲的异源免疫(heteroimmune)与同源免疫(homoimmune)的转导子都很有

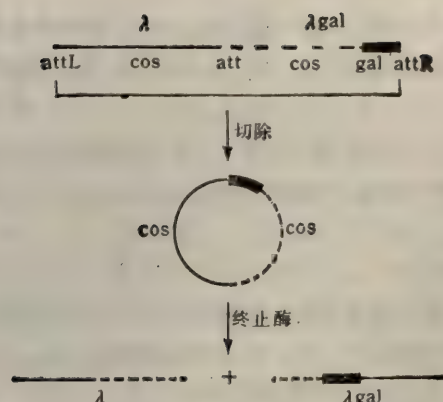


图 16-26 串联的双溶源化( $\lambda$ ,  $\lambda$ gal)细菌中产生  $\lambda$  与  $\lambda$ gal.

用,因为这些双溶源化细菌含有一个  $\lambda$ gal 和一个正常前噬菌体,它们带有有功能的头和尾基因。当进行诱导时,这些基因能提供  $\lambda$ gal 包装所需的结构组分。因此,当双溶源菌被诱导后,按图 16-26 所示的机理,可以产生  $\lambda$ gal 与正常的颗粒,两种颗粒的数目大致相等,由于  $\lambda$ gal 与正常  $\lambda$  的蛋白质含量都一样,而其 DNA 含量不同,所以可通过 CsCl 密度梯度离心将两种颗粒分离开来\*。

上述的双溶源化细菌裂解液中有一半是转导性颗粒。在实验室术语中,这样的裂解物被称作高频转导 (HFT) 裂解物。还有另一种裂解物,是由一个单个溶源菌形成的,由于发生错误的剪切,裂解物几乎没有或很少有一些转导颗粒,这种裂解物被称作低频转导裂解物 (LFT)。

## 参 考 文 献

- Campbell, A. 1976. "How viruses insert their DNA into the DNA of a host cell." *Scienc. Amer.*, December. pp. 102—113.
- Campbell, A. 1977. "Defective bacteriophages and incomplete prophages." *Comprehensive Virology*, 8, 259—328.
- Devoret, R. 1979. "Bacterial test for potential carcinogens." *Scienc. Amer.*, August, pp. 40—49.
- Ebel-Tsipis, J., D. Botstein, and M. Fox. 1972. "Generalized transduction by phage P22 in *Salmonella typhimurium*. I. Molecular origin of transducing DNA." *J. Mol. Biol.*, 71, 433—438.
- Hershey, A. D. (ed.) 1971. *The Bacteriophage Lambda*. Cold Spring Harbor Laboratory.
- Herskowitz, I., and D. Hagen. 1980. "The lysis-lysogeny decision of phage  $\lambda$ : explicit programming and responsiveness." *Ann. Rev. Genetics*, 14, 399—446.
- Howe, M. M. 1980. "The invertible G segment of phage Mu." *Cell*, 21, 605—606.
- Landy, A., and W. Ross. 1977. "Viral integration and excision: structure of the lambda att site." *Science*, 197, 1147—1160.
- Luria, S. E., J. E. Darnell, D. Baltimore, and A. Campbell. 1978. *General Virology*. Wiley.
- Miller, H. I., J. Abraham, M. Benedeck, A. Campbell, D. Court, H. Echols, R. Fischer, J. M. Galinda, G. Guarneros, T. Hernandez, D. Mascarenhas, C. Montaney, D. Schindler, U. Schmiessner, and L. Sosa. 1980. "Regulation of the integration-excision reaction by bacteriophage  $\lambda$ ." *Cold Spring Harbor Symp. Quant. Biol.*, 45, 439—446.
- Miller, H. I., A. Kikuchi, H. A. Nash, and R. A. Weissberg. 1979. "Site-specific recombination of bacteriophage  $\lambda$ : the role of host gene products." *Cold Spring Harbor Symp. Quant. Biol.*, 43, 1121—1126.
- Mizuuchi, M., and K. Mizuuchi. 1980. "Integrative recombination of bacteriophage  $\lambda$ : extent of the DNA sequen-

\* DNA 与蛋白质密度分别为 1.7 和 1.3 克/厘米<sup>3</sup>,噬菌体的密度决定于 DNA 与蛋白质的比例。



- ce involved in the attachment site function." *Proc. Nat. Acad. Sci.*, 77, 3220—3224.
- Nash, H. A. 1978. "Integration and excision of bacteriophage  $\lambda$ ." *Current Topics Microbiol. Immun.*, 78, 171—199.
- Nash, H. A. 1981. "Integration and excision of bacteriophage  $\lambda$ : the mechanism of conservative site-specific recombination." *Ann. Rev. Genetics*, 15, 143—168.
- Nash, H. A., K. Mizuuchi, L. W. Enquist, and R. A. Weissberg. 1980. "Strand exchange in  $\lambda$  integrative recombination: genetics, biochemistry, and models." *Cold Spring Harbor Symp. Quant. Biol.*, 45, 417—428.
- Oppenheim, A., and A. B. Oppenheim. 1978. "Regulation of the *int* gene of bacteriophage  $\lambda$ : activation by the *cII* and *cIII* gene products and the roles of the *pI* and *pL* promoters." *Molec. Gen. Genetics*, 163, 39—46.
- Ptashne, M., A. Jeffrey, A. D. Johnson, R. Maurer, B. T. Meyer, C. O. Pabo, T. M. Roberts, and R. T. Sauer. 1980. "How the  $\lambda$  repressor and  $Cro$  work." *Cell*, 19, 1—11.
- Weissberg, R. A., S. Gottesman, and M. E. Gottesman. 1977. "Bacteriophage  $\lambda$ : the lysogenic pathway." *Comprehensive Virology*, 8, 197—258.

## 第十七章 遗传物质重排与交换机理 (I)

### 质 粒

质粒(plasmid)是存在于多数细菌和一些真核生物染色体外的环状 DNA 分子。在一般情况下,对宿主细胞来说,质粒并非必不可少。例如,有时在细胞分裂时,产生不含质粒的子细胞,这个子细胞仍能生存下去。但是,许多质粒含有特殊条件下必须的基因。例如,R 质粒携带着许多种抵抗抗菌素的基因,为宿主抵抗多种抗生素提供可能,使宿主在实际中遇到无论是人为施给或是真菌产生的抗生素时,都能够生存下去。有一些质粒,例如 F 质粒的变种 F',含有一个或多个最初从宿主染色体转来的基因,如 leu 基因。这样的 F' 质粒对于在无亮氨酸的培养基中的 leu<sup>-</sup> 细胞的生长是必需的。这种质粒和其它许多质粒都是通过它们所携带的基因而被人们检测到的;然而,还另有一些质粒是当人们从细胞提取物中分离 DNA 样品时,样品中存在有小的环形 DNA 分子而被人们发现的。

本章将介绍质粒的许多有趣的生物学性质。同时,它们也是分子生物学工作者极其有用的工具,就像在第十四章讨论 lac 操纵子指出的那样,在遗传工程中极其有用(第二十章)。

### 质粒的遗传学性质及类型

虽然质粒有自己的基因和调节位点,而且不同质粒的基因和调节位点各不相同,但是就像噬菌体一样,它们的复制在很大程度上依靠宿主的细胞代谢功能。例如,质粒一般都使用宿主的大部分复制装备。然而,每种类型的质粒又都有它自己的调节质粒合成时间的基因以及调节每个细胞中质粒拷贝数的基因(严谨型或低拷贝数质粒在每细胞中有 1—2 个质粒;松弛型或高拷贝数质粒在每个细胞中有 10—100 个质粒)。而且,在细胞分裂时,质粒复制子(replicas)在子代宿主细胞中的分离分配也是受到严格调控的。也就是说,如果一个待分裂的细胞含有两个质粒 DNA 分子,那么每个子细胞分子就会分到一个质粒分子;如果出现了无质粒的细胞,就说明这个过程并未进行。这种事件的发生频率很低,每一个世代中,每 10<sup>4</sup> 个细胞中只有一个细胞会出现这种情况。这种稳定性来自质粒-染色体的协同联系。这种认识来自图 17-1 所示的实验,虽然这种协同联系的本质现在还不清楚。

如果没有特别说明,本章中我们所讨论的质粒就限于大肠杆菌的质粒。在大肠杆菌的各个变种中发现了许多种质粒,已研究过的有三种: F、R 与 Col 质粒。这些质粒有一些性质是相同的,但绝大多数性质很不相同。如果一个细胞有 F、R 或 Col 质粒,它就会出现下述性质:

1. F, 性质粒。这种质粒能够转移染色体基因(也就是说,可以转移并不是携带在质粒上的基因),并且能把 F 质粒本身转移到缺乏这种质粒的细胞中。



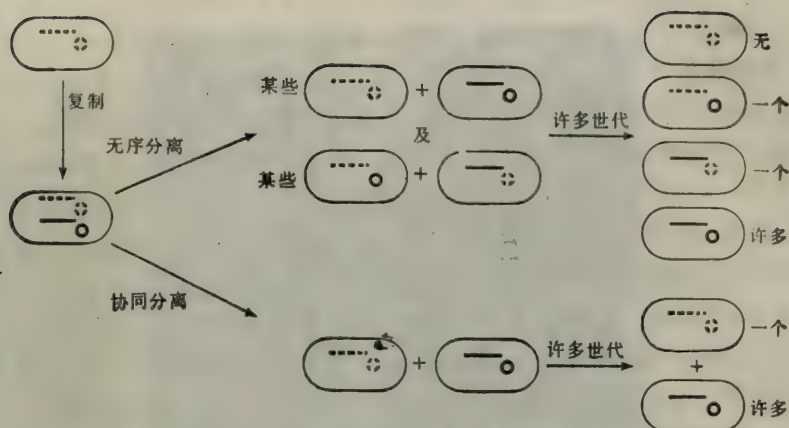


图 17-1 在这个实验中,细菌具有一个含放射性的染色体(虚线)和一个含放射性的质粒。此细菌再在不含放射性的生长培养基中生长很多世代。在协同的分离分配中,放射性全都存在于同时具有放射性的染色体和质粒的那个细胞中。观察结果如图所示。图中的“无”、“一个”、“许多”表示在那个世代中这种类型细胞的数目。

2. R, 抗药性质粒。这种质粒能抵抗一种或多种抗生素,并能给予无 R 质粒的细菌以抗药性。

3. Col 大肠杆菌素生成因子质粒。这种质粒能够合成大肠杆菌素 (colicins)。大肠杆菌素的这种蛋白质能够杀死与大肠杆菌亲缘关系很近但没有 Col 质粒的菌株。

本章将进一步讨论以上的每一种质粒。

已经知道质粒还带有一些其他基因。有一种在临床上很重要的质粒,带有一些能合成为肠毒素的肠内刺激物的基因。这些质粒称作 Ent, 它使旅行者发生腹泻以及某些痢疾疾病。另一种质粒使枯草杆菌(*Bacillus thuringiensis*)有合成毒素的能力,这种毒素对于舞毒蛾和幼虫有害。在较后面的段落中将不再介绍细菌中的质粒,而是介绍除细菌以外的一些生物有机体中的质粒。

## 质粒 DNA: 性质及分离方法

除了一个例外(酵母的杀伤质粒,是 RNA 分子),所有质粒都是超螺旋的环状 DNA 分子(图 17-2)。最小的质粒分子量为  $10^6$ ,最大的质粒分子量则稍大于  $10^8$ ,表 17-1 列出了几种研究较活跃的质粒的分子量。

在含有质粒的细菌中,经常发现质粒的二聚体和多聚体,这些形式是通过遗传重组形成的,较小的质粒更容易形成这样的形式。

质粒 DNA 一般可以用简单的方法从细菌中提取出来。向含有质粒的细菌培养物中加入去污剂使细菌裂解,然后离心裂解液。含有蛋白质与 RNA 的染色质复合体很大,而且很紧密,能较快地移向离心管的底部,质粒 DNA 则较小,留在上清液中。向清澈的裂解液的上清液加入 CsCl,离心直至裂解液的密度与 CsCl 密度相互平衡为止。在裂解液中通常掺有一些染色体 DNA,但如果质粒 DNA 的 (G + C) 含量与染色体 DNA 的



图 17-2 质粒 DNA 的一个开环形式与两个超螺旋形式,在超螺旋引入一个缺刻时,就产生开环形式。

表 17-1 几种质粒的一些性质

质粒	M $\times 10^6$	质粒拷贝数/染色 体	自我可转移性	表现型性质
大肠杆菌质粒				
ColE1	4.2	10—15	不	大肠杆菌素 E1(膜改变)
ColE2 (shigella)	5.0	10—15	不	大肠杆菌素 E2(DNase)
ColE3	5.0	10—15	不	大肠杆菌素 E3(ribosomal RNase)
性质粒				
F	62	1—2	是	F 纤毛
F'lac	95	1—2	是	F 纤毛; lac 操纵子
R 质粒				
R100	70	1—2	是	Cam <sup>+</sup> Str <sup>+</sup> Sul <sup>+</sup> Tet <sup>+</sup>
R64	78	(有限的)	是	Tet <sup>+</sup> Str <sup>+</sup>
R6K	25	12	是	Amp <sup>+</sup> Str <sup>+</sup>
PSC101	5.8	1—2	不	Tet <sup>+</sup>
噬菌体质粒				
$\lambda$ dv	4.2	$\approx 50$	不	$\lambda$ 基因 cro, cI, O, P
重组质粒				
PDM500	9.8	$\approx 20$	不	果蝇 melanogaster 的组蛋白基因
PBR322	2.9	$\approx 20$	不	高拷贝数
PBR345	6.7	$\approx 20$	不	ColE1 类型复制
酵母质粒				
2 $\mu$ m	4.0	$\approx 60$	不	没有已知的基因



(G + C) 含量不同,那么两种分子会形成不同的带,那就能够分别取出。但在通常的情况下,两者的 (G + C) 含量没有什么差别,因此,我们必须利用质粒是超螺旋化的这一事实。为了造成超螺旋与线型 DNA 分子在密度上的不同,在 CsCl 密度梯度离心中,加入溴化乙锭 (见第四章),超螺旋就具有较高的密度。这种广泛应用的技术的唯一缺点是,在分离中偶然形成的带缺刻的环状 DNA 和非超螺旋型的复制型的分子会随着染色体 DNA 一起被弃去。

通过电泳以一种极理想的方式可检测出一个单个的细菌克隆中的质粒 DNA (图 17-3)。取一个细菌克隆,将细菌裂解,然后进行电泳。由于细菌的染色体 DNA 很大不能进入凝胶(虽然它的一些片段能进入),但质粒 DNA 能够进入凝胶。DNA 分子通过凝胶的电泳速度随 DNA 的分子量减小而加大,因此如果含有质粒 DNA,它就会在凝胶中反映其分子量的特定位置上形成一条很窄的带。通过溴化乙锭染色,可以见到这个条带。溴化乙锭与 DNA 紧密结合,在紫外线照射下能发出荧光。如图所示,从一段时间内未知的质粒 DNA 移动距离与一个已知分子量的质粒所移动的距离进行比较,可以计算出质粒 DNA 的分子量。以上筛选方法在重组 DNA 技术中非常有用,读者将于第二十章见到。用电泳技术也可从清澈的裂解液(见上)中检测到质粒 DNA。

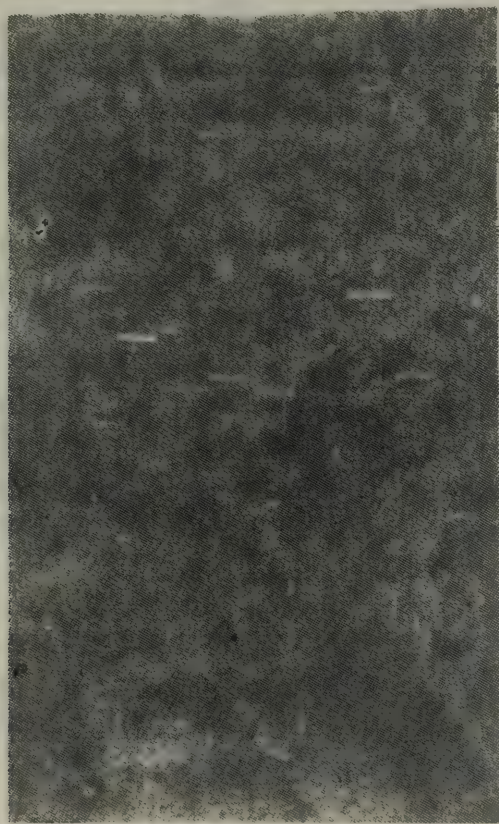


图 17-3 质粒的凝胶电泳表示出 8 种不同质粒的迁移。泳动从上到下进行,每一个竖行“泳道”代表一种质粒,电泳后将凝胶浸入溴化乙锭溶液。待漂洗后,再用近紫外线照射,与 DNA 结合的溴化乙锭发出荧光。在每条泳道中的单个亮带就是超螺旋 DNA 分子,移动得较慢的暗些的条带是缺刻环或超螺旋的二聚体,电泳的装置见图 3-14(承 Elaine Locuzzo 和 Pieter Wensink 提供)。

分离 DNA 通常还必须有除去蛋白质的步骤,如果不用除蛋白质步骤,那么从某些大肠杆菌用上述方法分离 DNA 分子,大约有一半的超螺旋质粒 DNA 分子上附着着三种紧密结合的蛋白质分子。这个 DNA-蛋白称作松弛型复合体 (relaxation complex)。如果对这个复合体加热或用碱处理,或者用蛋白酶处理或者用去污剂处理,留下有核酸酶活性的蛋白质,它在 DNA 的一条链上形成缺刻,从而将超螺旋放松并对环状 DNA 形成缺刻(图 17-4),这种缺刻只发生在一条链上并只发生在一个专一的位点上。在松弛过程中,释放出两个较小的蛋白质,最大的那个蛋白质则共价连接到缺刻的 5'-P 末端上。如果在松弛之前用一些实验室技术使超螺旋 DNA 形成缺刻,那么上述引起松弛的核酸酶就不再能形成上面提到的位点专一的缺刻。这说明这种松弛型核酸酶只作用于超螺旋 DNA。人们认为这种缺刻在质粒通过接合转移而转到另一个细菌的过程中起作用(参见下节)。

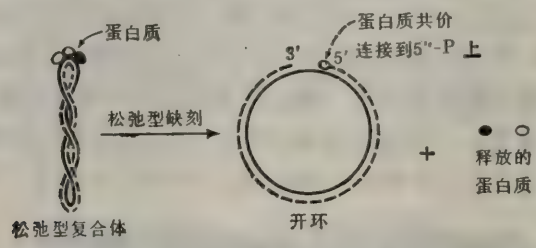


图 17-4 在超螺旋 DNA 松弛性复合体的一条链上缺刻的形成。

## 质粒 DNA 的转移

许多种质粒能将质粒复制子从供体细菌转移到受体细菌中。例如,如果将含有 F 质粒的大肠杆菌加到正在生长的无 F 的细胞培养液中,在细菌生长了许多世代之后,大部分细胞都含有 F。这是 F 复制子从 F<sup>+</sup> 细胞中以不丢失 F 的方式转移到 F<sup>-</sup> 细胞的结果。转移之后,原来细胞中的 F 质粒仍能复制,复制物可再转入新的 F<sup>-</sup> 细胞。每个获得 F 质粒的受体细胞又成了供体,可对 F<sup>-</sup> 细胞供给 F 质粒。每个世代可发生一至两次转移,所以 F 很快地在细菌群体中扩散开来。对于质粒来说,其接合能力是它达到成功的进化的策略之一。在细菌接合期间,宿主的 DNA 不能复制,质粒基因能复制;这种质粒的复制过程为质粒的散布提供了明显的有利条件。质粒通过转移增强这种散布过程并在许多种宿主中稳定地遗传着。

### 转移的大致步骤及某些定义

质粒转移过程可分为四个时期:

1. 形成专一性的供体-受体对(又称为有效接触, effective contact)。
2. 准备 DNA 转移(又称为移动, mobilization)。
3. DNA 进行转移。
4. 在受体中形成有复制功能的质粒(又称为形成复制, repliconation)。

许多质粒不能实现上述全部过程,所以又发展出下面几个术语:

接合性质粒 (conjugation plasmid): 带有决定有效接触功能基因的质粒。



可动质粒 (mobilizable plasmid): 有可能将其 DNA 进行移动的质粒。

自身转移质粒 (self-transmissible plasmid): 例如 F 质粒, 兼有可接合性及可动性的质粒。

其他一些质粒如 ColE1, 是可动的但不是接合性质粒。自然界中不存在是接合性的但又是不可动性的质粒。但是, 突变了质粒则有可能具有这种性质。

一个细胞中可以含有两种不同的质粒, 如 F 质粒与 ColE1 质粒。因为 F 既是接合性的, 又是可动的 (系自转质粒)。它可为 ColE1 质粒提供接合功能, ColE1 本身是可动的, 于是 ColE1 也就能被转移了。这个过程, 即接合性质粒所形成的有效接触可以帮助非接合性质粒, 使之得到转移, 称作供体作用 (donation)。

供体作用的特征就是非接合性质粒的有效转移。此外, 一个接合性质粒同样也能帮助一个非可动质粒得到转移。这个过程叫做转导作用 (conduction), 以区别于供体作用。转导作用过程发生的频率很低, 它是由两个质粒之间发生基因重组形成一个可转移的 DNA 的单个分子而完成的。一个非自转质粒受帮助后的转移频率 (高对低) 点是判断一个质粒通过供体作用或是通过转导作用是否是一个可动质粒的有用标准。

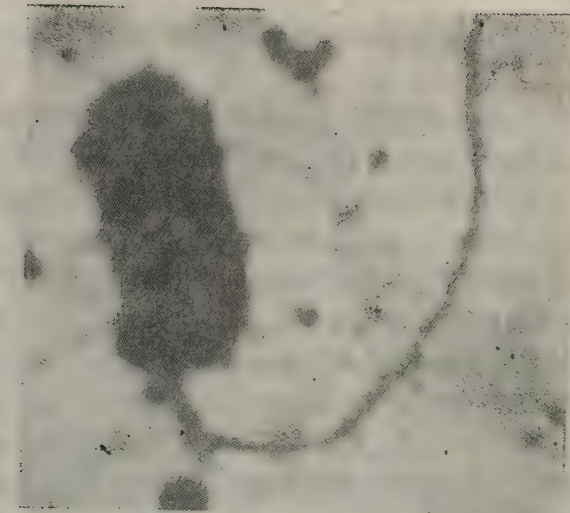
F 质粒的分子量为  $63 \times 10^6$ , 其中有转移所需要的至少 19 个基因, 这些都称为 tra 基因。这些基因中的任何一个基因发生突变都会使质粒失去自身转移能力, 因此人们推测对于任何自身转移质粒来说, 自身转移过程所必需的基因至少是 19 个。因为这些基因需要的分子量约为  $15 \times 10^6$  分子量单位的 DNA (假设基因产物的平均分子量为  $4 \times 10^4$ ), 因此小的质粒不是自身转移质粒, 已知最小的自身转移质粒的分子量为  $17 \times 10^6$ , F 质粒将作为研究转移过程的模型在下面加以介绍。

## F 质粒转移的细节

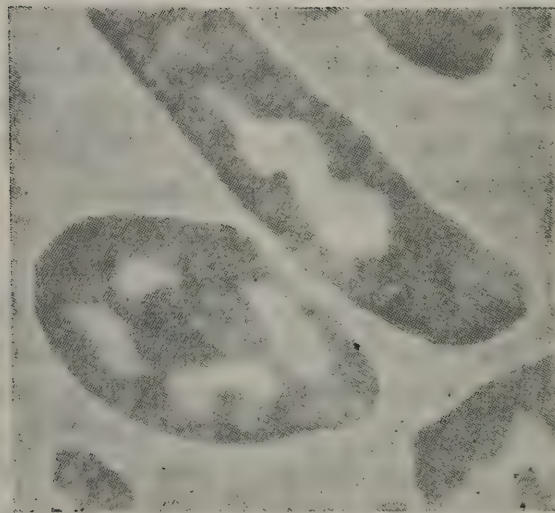
有效接触的第一步是供体与受体配对, 这要求供体细胞上有称作性纤毛的头发样蛋白质附件 (图 17-5), 含有 F 或 R 质粒的细胞的纤毛分别称作 F 纤毛 (F pili) 或 R 纤毛 (R pili)。纤毛是一个空管, 由称为纤毛素 (pilin) 的一种蛋白质构成, 其顶端与受体细胞接触。合成纤毛所需的十二个基因中只有一个是用于合成纤毛素的。人们已经假设 DNA 是通过纤毛而实现转移的, 但还没有很好的证据支持这种看法。在配对之后, 纤毛又缩进供体细胞。因此, 纤毛很可能首先是使两个配对的细胞相互初步地接触, 然后再使两个细胞紧密地接触。

当质粒所编码的一种蛋白质在单链上的一个称之为转移起始区 (transfer origin 或 Ori T) 的特殊序列处形成一个缺刻时, 转移过程就开始了。这种蛋白质很可能是松弛复合体的缺刻蛋白, 还没有在所有的质粒中检测到松弛复合体, 但是推测有一种蛋白质执行着上述功能。形成缺刻, 起始着滚动环复制, 接着, 滚动环复制产生的线性 DNA 分枝再被转移。人们认为形成缺刻的蛋白质与 5' 末端结合, 质粒复制的方式与  $\phi$ X174 噬菌体的带泡的滚动环复制方式相同 (见第八章图 8-40), 图 17-6 表示了转移顺序过程中各个事件的发生顺序。

在供体和受体细胞中都进行 DNA 合成。在供体中称为供体的接合 DNA 合成 (donor conjugal DNA synthesis), 其目的是代替被转移走的单链。在受体中称为受体的接合 DNA 合成 (recipient conjugal DNA synthesis), 是将转移来的单链变为双链分



(a)



(b)

图 17-5 (a) 具有一个性纤毛的大肠杆菌细胞。性纤毛上吸附有专一于 F 的噬菌体 R17, R17 使其细的纤毛附着一层粗黑色吸附物而可以观察到。五条很亮的纤维是鞭毛, 不清晰的非常细的是纤毛(承 Barry Eisenstein 提供)。(b) 二个大肠杆菌细胞结合时的电镜照片, 小的是 F 细胞, 大的是有 F'<sup>lac</sup> 的细胞(承 Lucien Caro 提供)。

子。DNA 以线状分子的形式进行转移。但形成的质粒 DNA 是环状的。所以需要重新环化的过程, 但这个过程还不清楚。

通常情况下, 在转移期间, 被转移的单链被供体接合合成的 DNA 取代; 同时, 通过受体的接合 DNA 合成, 该被转移的单链在受体细胞中形成双链 DNA 分子。下列现象可说明 DNA 合成与转移并不偶联到一起: (1) 供体接合 DNA 合成被适当的宿主突变所抑制时, 转移仍可进行; (2) 当转移被适当的质粒突变作用所阻止时, 供体接合 DNA 合



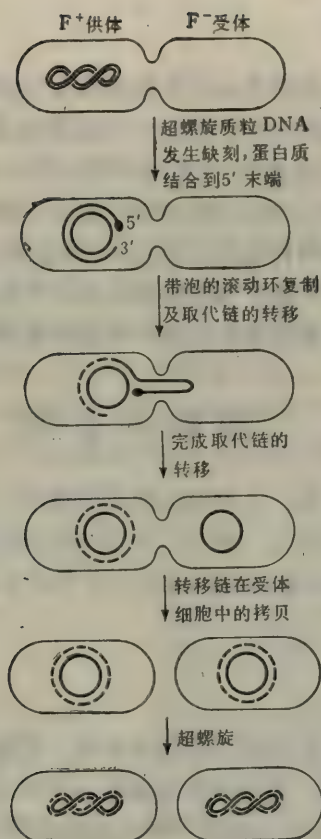


图 17-6 通过滚动环机理从 F<sup>+</sup> 细胞转移 F 质粒 DNA 的模型。被置换的单链, 转移到 F<sup>-</sup> 受体细胞并在那里转变成双链 DNA。图中已将染色体 DNA 省略。

成也仍可进行;(3)由于受体细胞突变使受体接合 DNA 合成不能进行时,转移也仍进行。这些现象清楚地说明 DNA 复制不是转移的动力。由于还有双链螺旋 DNA 分子解旋的驱动力的问题,所以有关转移的动力学问题是十分微妙复杂的。在第八章中强调了 DNA 聚合酶 III 不能使双螺旋解链。所以,在没有其他因子的帮助下,DNA 聚合酶 III 不能单独地复制双链 DNA 分子。解链需要一种辅助蛋白——解链酶(helicase)。因此,当转移进行时,解链酶必须活化,以便供体 DNA 解链,使转移链可以转移。在解链酶无活性时,复制不能进行。但不进行复制时,解链酶也可发挥作用。因此,有人提出转移起始时需要二个条件——转移起始点有一个缺口以及解链酶的活化。活化后解链酶将供体 DNA 分子解链,于是单链可以转移。解链酶的解链作用将引起质粒的超螺旋程度的改变,在第八章中已指出由此引起的这种几何紧张状态可在 DNA 解旋酶 (gyrase) 的作用下缓解。所以可以预期,当解链酶或解旋酶的活性受到抑制时,转移不能进行。萘啶酮酸 (nalidixic acid) 是一种大肠杆菌 DNA 解旋酶的抑制剂,当把它加到细菌培养基中,时转移就不能进行,这个现象可以支持(但不能证实)上述假设。遗憾的是,目前还没有发现解链酶的抑制剂。

## 通过质粒 DNA 的转化作用

上节所描述的质粒 DNA 的自然转移是通过细胞与细胞的接触进行的。在实验室中则通过将游离的质粒 DNA 与经过适当处理的受体细胞混合来实现转移,即转化。如果将细菌用经过预冷的  $\text{CaCl}_2$  溶液处理,它们将具有接受外来 DNA 分子的能力。这就为将任何可分离的质粒转移到任一大肠杆菌菌株中提供了可能性。这种转移的两个条件是:(1)受体菌的复制酶体系同样适用于质粒。当受体与供体是同一种系时,总是如此。(2)受体与供体细胞的宿主修饰-限制体系(第十五章)要相同。目前, $\text{CaCl}_2$ -转化作用是将不能传播的质粒从一种菌株转移到另一种菌株的最普遍和最重要的实验技术。

## 质粒的复制

质粒只能在宿主细胞中复制。人们可以由此而猜想,天然地生长在同样的宿主细胞中的所有质粒都有相同的复制类型。但是,与那一些仅仅能在同一种宿主细胞内复制的噬菌体一样,各种质粒的 DNA 复制在其酶学上及机理上是各种各样的。这种特点将在下面两节中予以解释。

### 对宿主蛋白质利用的差异

质粒的复制在很大程度上依赖宿主的复制系统。实际上,有些质粒似乎是完全地靠着宿主的基因产物来进行复制。如在体外将纯的 ColE1 DNA 加入到不含 ColE1 及其他质粒,即不含有任何质粒所编码的基因产物的细胞提取物中,小质粒 ColE1 可以在这个体外体系中进行复制,(这个结果表明:在没有质粒蛋白质存在的条件下,ColE1 DNA 是可以复制的。当然我们也不排除在体内复制时利用质粒蛋白质的可能性)。而另外一些质粒则还是需要某些质粒所编码的基因产物。例如,已知 F 质粒的温度敏感突变体( $F^-$ )在  $42^\circ\text{C}$  不能复制,但在此温度下细菌 DNA 仍可正常地合成。把含有温度敏感的  $F^-$  质粒的细胞放在  $42^\circ\text{C}$  培养几代之后,将得到有  $F^-$  的子代细胞。许多质粒的复制过程并不需要全部的宿主基因产物。这既可能是由于质粒的基因能编码出一种与宿主中存在的蛋白质相类似的基因产物,又可能是存在着一种截然不同的复制模式。例如, F 可在大肠杆菌  $\text{dna A(Ts)}$  突变株中,于  $42^\circ\text{C}$  的高温下进行复制。而此时细菌的染色体 DNA 不能复制。所以, F 的复制并不需要大肠杆菌的  $\text{dnaA}$  基因。

此外,还有一些质粒,它们所需要的宿主产物对宿主来说也是有用的,但不是那么严格需要。例如,在某些细菌的突变株中, DNA 聚合酶 I 的活性很低,但细菌仍可正常地生长,而在这种细菌突变株中,ColE1 质粒不能进行 DNA 复制。

### 复制机理的差别

所有质粒已被证实为半保留复制,并且在复制周期内保持环状。可是不同的质粒的复制类型如同我们已知的噬菌体体系一样,存在着显著的差别。

下面所列的就是一些不同点:

1. 方向性: 现已观察到,有的质粒只有单向复制,有的质粒只有双向复制;此外,也见



到有两种复制类型的质粒。例如 PK6 质粒,首先沿一个方向复制,一段时间后,又在同一起点沿相反方向复制。

2. 终止: 单向复制的质粒经一个复制周期后,其复制一定终止在起点处。双向复制的终止点已观察到有两种类型: 一种是当生长着的两个复制叉相遇到达同一区域时,复制终止。如  $\lambda$  噬菌体(第八章)。另一种是有固定的终止位点。有时,一个复制叉比另一个复制叉先到达该点。目前,尚未弄清楚 DNA 复制的终止信号究竟是什么。

3. 复制类型: 对质粒的认真研究发现质粒的复制是所谓蝴蝶型的,这种蝴蝶型方式首先在动物病毒中观察到(图 17-7)。在部分复制的分子中,复制的那些部分是不扭旋的,即是通常所说的  $\theta$  型复制,但不复制的那些部分则是超螺旋化的。当复制周期结束时两个环状分子中的一个一定被切开(DNA 解旋酶可能起作用)。于是,一个复制周期产生出一个有缺刻的分子和一个超螺旋的分子。有缺刻的分子,经过一段时间后被补上缺口。并且,再过一段时间后又变为超螺旋状态。目前还不清楚这是否就是质粒复制的一般机理。

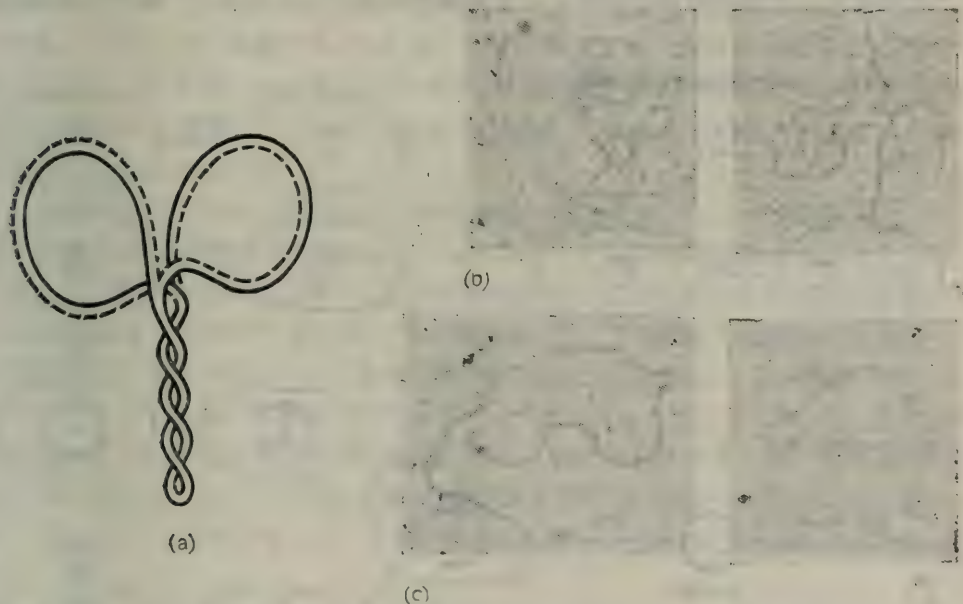


图 17-7 ColE1 DNA 复制。(a) 蝴蝶型分子示意图,虚线是新合成链。(b) 蝴蝶型分子的电镜照片。(c) 有缺刻的蝴蝶型分子,缺刻使蝴蝶型分子变成  $\theta$  型分子(承 Donald Helinski 提供)。

### 拷贝数的控制

有些质粒在细胞中是低拷贝的——每个细胞中有一个或几个。同时,也存在其他一些高拷贝数的质粒,每个细胞中有 10—100 个质粒。

下面通过简单的实验可以说明拷贝数的确立和调节是受 DNA 合成起始速度控制的。

如果将单拷贝质粒或低拷贝质粒转化到细胞中,在细胞分裂之前,质粒 DNA 只复制一次或二次。而细胞若经转化得到一种高拷贝数的质粒,这种质粒 DNA 将反复地复

制,直至达到合适的拷贝数为止。

对此现象,人们所能接受的解释是:质粒所编码的阻遏物结合到质粒的操纵基因上,抑制复制,从而实现对其拷贝数的调控。首先,让我们利用这个想法来解释质粒的拷贝数从上一代到下一代是如何保持稳定的。目前的理论,假定阻遏物只有在多个亚基形成完整的蛋白质时才有活性,单体与多体之间的平衡紧密地依赖于阻遏物的浓度。因而当细胞生长时,阻遏物浓度下降,变为无活性的单体形式;复制作用的抑制因此而解除,质粒 DNA 分子数成倍地增加。此时,阻遏物基因的数目也变为开始时的二倍。接着,通过蛋白质生物合成也使基因产物——阻遏物的浓度由此加倍。结果形成有活性的阻遏物,于是质粒的复制终止。如果起始时只有一个高拷贝型的质粒,情况也类似:复制继续不断地进行,直到有足够的阻遏物,质粒 DNA 的复制才终止。这种看法又是如何解释拷贝数的不同呢?最大的可能性是,对于高拷贝数的质粒来说,需要有比低拷贝数质粒更多的有活性的阻遏物,才能使复制终止。因此,相对于低拷贝数质粒,它所需要的单体阻遏物浓度要更高一些,以便缔合出更多的、有活性的多体阻遏物分子。因此,只要当每个细胞内的质粒的拷贝数高到可以产生出足够量的有活性的阻遏物时,即“基因剂量”(gene dosage)高达一定程度时,质粒的复制才能终止。下面的实验支持以上看法——即每种质粒控制自

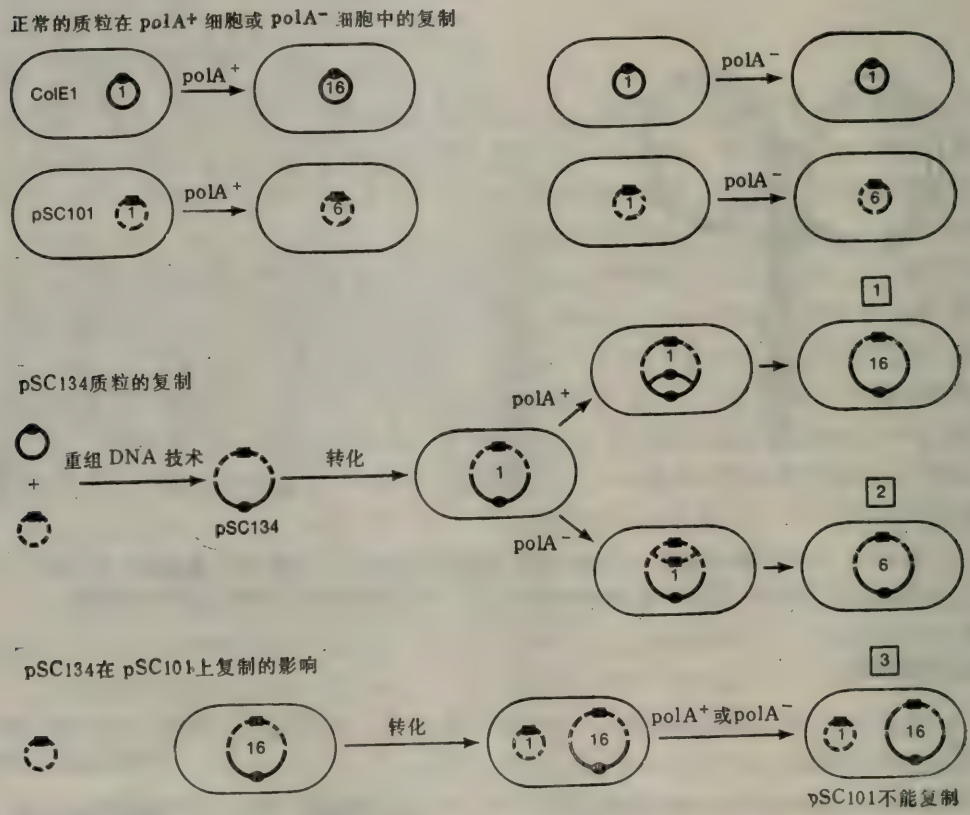


图 17-8 ColE1, pSC101 及杂交质粒 pSC134 复制示意图。开始时每个细胞中有一个拷贝,质粒环上的方块及圆点代表 pSC101 和 ColE1 的复制起点。DNA 分子内的数字 (1, 6 和 16) 代表每个细胞内的拷贝数,实线框中的数字 (1, 2 和 3) 代表文中的三个实验现象。



己的拷贝数。

杂交质粒 pSC134 是 ColE1 质粒和 pSC101 质粒通过重组 DNA 技术而构建的。这两种质粒的拷贝数分别是 18 和 6。以下三点事实是从 pSC134 的实验获得的(图 17-8):

1. pSC134 质粒,从 ColE1 的复制起点开始复制。其拷贝数是 16,大致等于 ColE1 的拷贝数。

2. ColE1 质粒在  $polA^-$  细胞中不能复制。如果把 pSC134 质粒放到  $polA^-$  细胞中,则质粒 pSC101 质粒的起点就被启用。pSC134 的拷贝数变为 6。与 pSC101 相同。这二个结果表明拷贝数目与所用的复制起点相关联。

3. 如果 pSC101 DNA 进入到具有 16 个拷贝 pSC134 的细菌中,pSC101 不能复制,表明 pSC134 产生出 pSC101 复制的阻遏物。

以上三个结果表明:如果在细胞中 pSC134 的拷贝数比 6 少,pSC101 和 ColE1 的起点都有活性,从两个起点开始的复制则使拷贝数增加。如果拷贝数比 6 多,pSC101 的起点失活,因为对 pSC101 复制起限制作用的阻遏物的浓度已高于抑制水平。虽然假若还有自动调节,pSC101 DNA 还可继续合成,可是其阻遏物的浓度也决不会超过 6 个拷贝所产生的量。ColE1 起点保持活性,只有当拷贝数比正常值多 1 的时候才关闭。在  $polA^-$  细胞中,ColE1 不能复制,拷贝数完全由 pSC101 的阻遏物浓度控制。所以,拷贝数不会超过 pSC101 的正常值。这完全可以解释以上现象。同时,还有其他一些实验也支持由阻遏物调节质粒拷贝数的模型。

接下去我们必须对阻遏物模型的重点有所理解,因为所有存在于一个特定细胞内的质粒是相同的,阻遏物不能分辨出不同的分子,因此,当活性阻遏物的浓度低到不足以使所有 DNA 分子的复制都受到抑制时,阻遏物将随机地只使其中一些分子的复制受到抑制。这就意味着,当一个质粒复制成两个子代质粒时,每一个单独的子代质粒与其他分子(第一次复制)有着同样的机率再进行复制(第二次复制)。这就是说,在任何情况下,全部质粒的每一个分子可以进行复制的机会是随机的。这个假设已被含有  $^{14}N$  标记质粒的细菌生长在含有  $^{15}N$  标记的培养基中的情况所证实。如果允许这种质粒的群体进行几次复制循环,那么在所有  $^{14}N^{14}N$  标记的质粒都转变成成为  $^{14}N^{15}N$  状态的杂合子(只进行了一次复制)之前,就可以观察到有一些  $^{15}N^{15}N$  标记的质粒(这些质粒已经经过两个复制循环)。

## 不相容性

通常一对对密切相关的质粒是不能同时地存在于一个单个细胞的子代细胞中的。人们称这些质粒是“不相容的”(incompatible)。上述阻遏物抑制质粒复制的模型,可以解释这种现象。

首先考虑一个细胞内同时存在着两种具有不同阻遏物的质粒(如 F 和 ColE1)的情况。由于一种质粒(如 F)产生的阻遏物不能调节另一种质粒(如 ColE1)的复制,所以它们的复制是彼此独立的,是相容的,因而说它们属于不同的“不相容性组群”(incompatibility groups)。

当存在着两种具有相同或相似阻遏物的质粒 A 和 B 时,阻遏和调节复制的情况将完

全不同。A 可以调节 B 的复制，反之亦然。让我们看一看一个细胞内有一个 A 和一个 B 的情况，这时细胞已经生长得足够大，正在发生着脱阻遏作用（derepress）（图 17-9）。由于二种质粒是随机复制的，所以第一次复制后，细胞中可能有 1 个 A 和 2 个 B（1A,2B）或 2 个 A 1 个 B（2A,1B）。第二次复制后每个细胞中将有 4 个质粒。由于这次被复制的质粒不同，所以质粒的成分可能有几种，如图所示；（1A,3B），（2A,2B）或（3A,1B）。此时，具有原来质粒数两倍的细胞可以进行分裂。两个子代细胞中所具有质粒的情况将如下：

- （1A,3B）变为（1A,1B）+（1B,1B）。
- （2A,2B）变为（1A,1A）+（1B,1B）或（1A,1B）+（1A,1B）。
- （3A,1B）变为（1A,1A）+（1A,1B）。

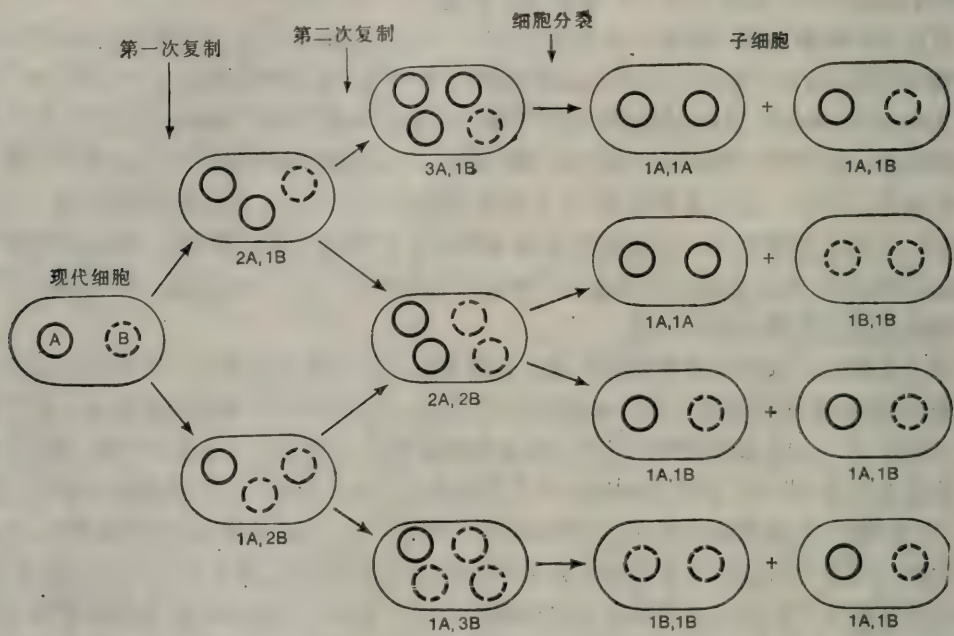


图 17-9 含有不相容质粒 A,B 各一个的细胞。分裂后,子代细胞含有质粒的四种可能情况。详见正文。

注意：（1A,1A）和（1B,1B）这两种细胞只含有一种质粒，它们的子代细胞当然也只有一种质粒。含有 1 个 A 和 1 个 B 的细胞继续分裂。子代细胞有 50% 可能失去一种质粒。因而，含有二个不相容性质粒的细胞，随着细胞分裂的不断进行，只含有一种质粒的子代群体的百分数将逐步增加。

因此，不相容性是由两种质粒具有类似的阻遏物和质粒 DNA 随机地复制两个因素导致的结果。

### 复制的抑制

质粒的复制受多种螯合剂的抑制，特别是吖啶类（如原黄素 proflavin；吖啶橙 acridine orange）。但是它们并不抑制染色体 DNA 的复制。这种抑制可能导致质粒的丢失（吖啶疗法，acridine curing）。这种现象在质粒含有宿主基因时，很容易发现，由图 17-10 说



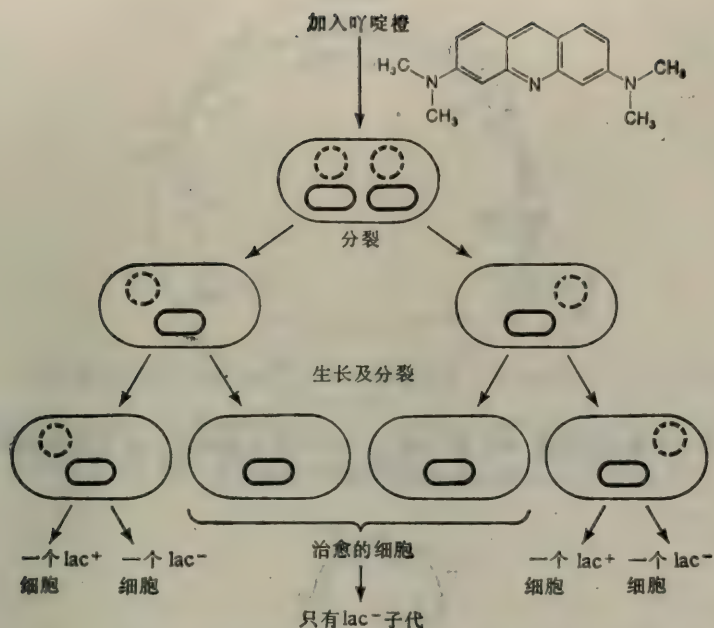


图 17-10 在含有四环素的培养基中,具有 F' lac<sup>+</sup> (圆形环)的细胞的治疗过程。

明。如果一个细菌培养物的基因型是 F' lac<sup>+</sup>/lac<sup>-</sup> (即染色体是 lac<sup>-</sup>, 同时 F 因子上带有 lac<sup>+</sup> 等位基因), 把它放在含有四环素的培养基中生长, 细菌可以不断地生长增殖。然而经过一个细胞周期后, Lac<sup>+</sup> 的细胞数保持不变, 而 Lac<sup>-</sup> 的细胞数则增加。这是由于加入四环素时, 几乎所有的细胞都具有两个拷贝的 F' lac。一个细胞周期后, 每个细胞只含有一个 F' lac, 由于质粒的复制被抑制, 因而随着下一次细胞分裂, 二个子代细胞中只有一个含有质粒, 这样就产生不含质粒的细胞。

## 特殊细菌质粒的性质

在这一节将详细讨论 F、R 和 Col 质粒以及其他几种质粒的性质。

### 性质粒 (F 质粒)

F 是一环状的 DNA 分子, 分子量为  $62.5 \times 10^6$  (94.5 kb)。它已被广泛地进行过遗传学和物理学图谱的分析。令人感兴趣的是 IS2, IS3 和  $r\delta$  序列示于图 17-11。这些区域叫转座子 (transposons), 将在本章及第十九章进一步讨论。

F 的一个重要性质是, 它具有整合进入细菌染色体从而产生 Hfr (第一章) 细胞的能力。这种整合与  $\lambda$  噬菌体形成溶源菌非常相似 (参见图 16-7)。但它整合进入大肠杆菌染色体与  $\lambda$  前噬菌体有以下几点不同: (1) 在 F 中, 整合不只在—个位点进行。整合主要在物理图谱中 94.5 附近的 IS3 处进行; 另外, 整合也可在其他的 IS3 和  $r\delta$  上发生, 如图 17-12 所示。(2) 在染色体上现已知道在 20 个以上的主要位点上都可发生整合 (图 17-13)。F 对染色体上的多个整合位点的亲和力是不同的, 所以由此去形成各个特定 Hfr 株的频率也是不同的。人们认为这些位点是 IS2, IS3 和  $r\delta$  在染色体上的拷贝, 所

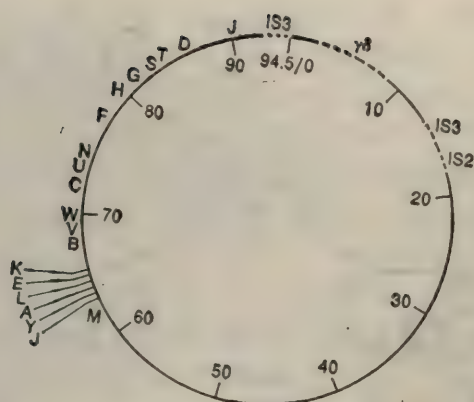


图 17-11 F 性质粒的图谱。单位是千碱基。大写字母是对应的各个 *tra* 基因的中点位置。虚线是  $r\delta$ , IS2 和 IS3 序列。

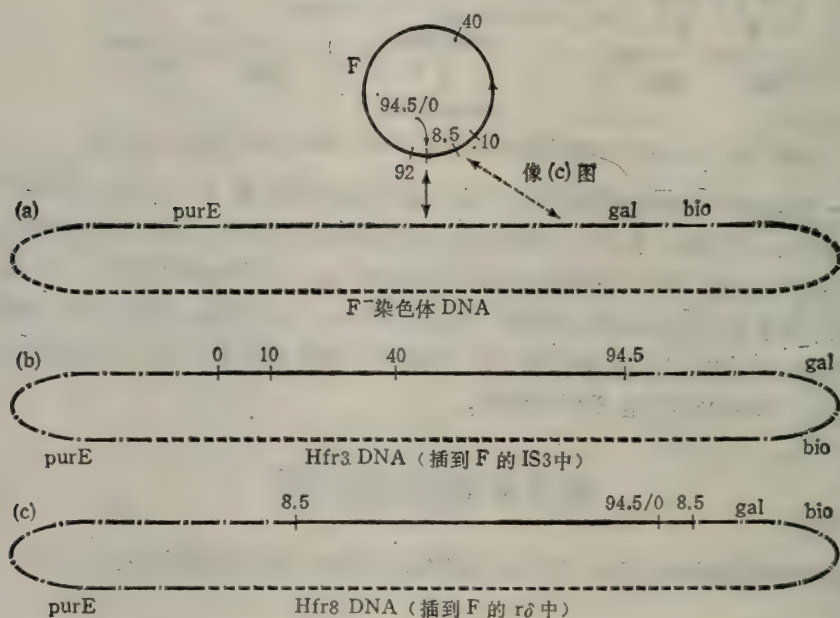


图 17-12 两种不同 Hfr 株的形成。(a) F<sup>-</sup> 染色体和 F 质粒。交换发生在实线双箭头所指的 两端。(b) (a) 中的交换结果。(c) (a) 中, 虚线双箭头两端发生交换的结果。

以有进行交换的可能。亲和性则代表着 F 中的 IS<sub>2</sub>, IS<sub>3</sub>,  $r\delta$  与染色体中的这些位点片段进行交换的概率。(3) 整合方向: F 的整合既可以是顺时针方向的, 也可以是逆时针方向的(见图 17-13)。F 到雌性细胞的转移是从一个单一位点沿单一方向进行的; 而有一些 Hfr 株转移染色体却是沿两个方向的。事实上, 有少数几个例子, 其 F 质粒可在同一个位点上沿两个方向进行整合。

F 质粒整合进入细菌的某些 *dna*<sup>-</sup> 突变株所引起的现象叫做整合性抑制(integrative suppression)。前面我们已提到 F 不需要大肠杆菌的 *dnaA* 基因产物, 而这个基因产物却是染色体复制起始所必需的蛋白质。如果 *dnaA*(Ts) (即 *dnaA* 的温度敏感突变株) 中



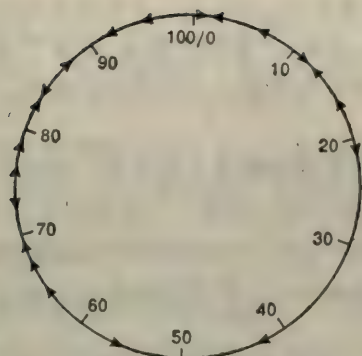


图 17-13 大肠杆菌的遗传学图谱。表示出 Hfr 转移的 20 个不同转移起点。箭头指出转移的方向。

含有一个游离的 F 质粒，当温度升高到 42℃ 时染色体不可能起始复制作用。而 F 仍然可以复制。与此相对，在 Hfr 菌株中 F 是整合的；同样地，这个菌也是一个 dnaA(Ts)突

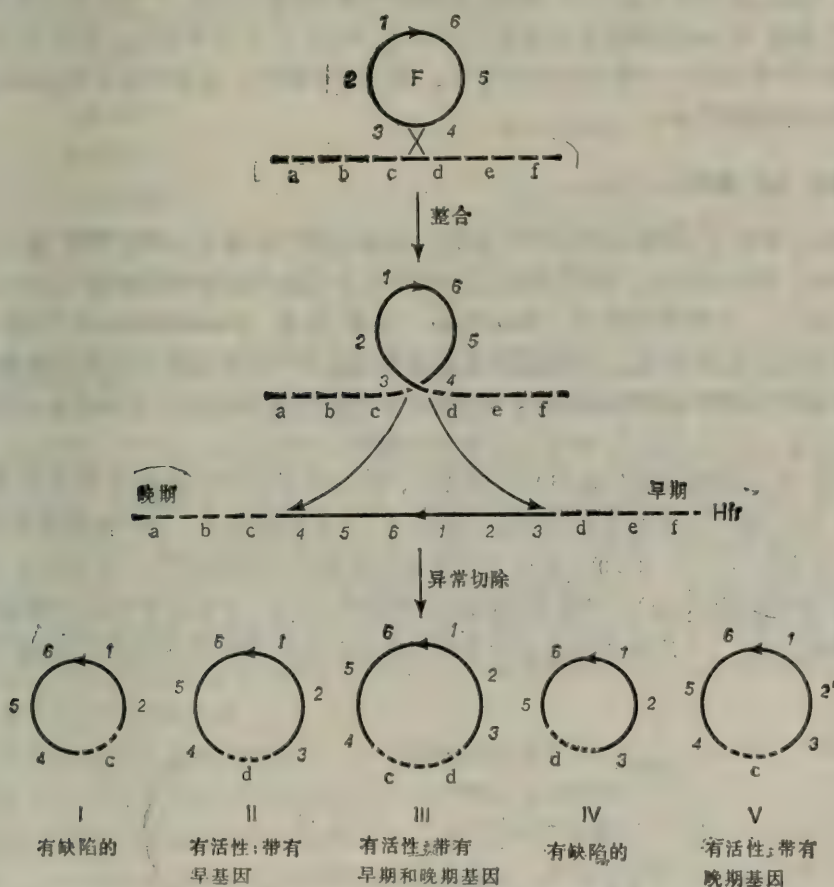


图 17-14 从特定的 Hfr 株中经异常切除所产生的不同 F 质粒的情况。I 及 IV 质粒失去了 F 基因，所以是有缺陷的。如果质粒是复制缺陷型的，它将不能被保存下来，也就无法观察到。通常 F' 质粒是通过它所带的基因而被检测到的，一般情况下，Hfr 细胞只转移晚期基因，而不转移早期基因。所以，I 和 IV 通常是观察不到的，因为它们是有缺陷的。同样，也观察不到 II 型质粒，因为它带有早期基因。

变株,在高温时染色体可以无限地复制,但复制不是在染色体的复制起始点处起始,而是在F的 *oriT* 位点处开始复制。因此,F质粒的整合,抑制了表现型 *DnaA(Ts)* (非常有趣的是,在低温时复制是由通常的染色体上的起始点起始的。至于为什么此时F的复制起始点不活化,还不清楚)。

F的切除偶尔也会发生。通常的切除是不完善的。切除F有二个切点,一个切点在整合的F DNA 的两个末端中的一端,而另一个切点则在相邻的染色体DNA中(图17-14)。这个过程与从 $\lambda$ 溶源菌产生专一性转导颗粒的过程相似(第十六章)。含有染色体基因的F叫做F'质粒(F' plasmids)。F'质粒的命名及遗传学性质在第一章和第十四章中已有叙述。

F'质粒是分子生物学研究中非常有用的工具。下列几种情况表明它的意义:(1)可用于带有局部双倍体的遗传学显性研究。例如用于对*lac*操纵子的分析(见第十四章)。(2)在发现重组DNA技术(第二十章)以前,提供了一个供给特定的染色体基因的方法。甚至还可以通过某些专门技术作为分离特定基因的手段。(3)用于研究前噬菌体的整合。例如,最初有关线性 $\lambda$  DNA整合的现象,就是在具有F' *att $\lambda$* 的溶源菌中观察到的。它表明F' *att $\lambda$* 的分子量是由于一个 $\lambda$  DNA分子而增加的。同时具有F' *att $\lambda$* 的前噬菌体分子仍为一个连续的环状分子。在这个实验中,也观察到了具有两个相邻的前噬菌体的双溶源现象。

## 抗药性质粒(R质粒)

抗药性(或称R)质粒首次是在日本的一次痢疾流行性爆发中在志贺杆菌(*Shigella dysenteriae*)里被发现的。以后,又在大肠杆菌以及其他种属如沙门氏菌(*Salmonella*)、弧菌(*Vibrio*)、枯草芽孢杆菌(*Bacillus*)、假单胞菌(*Pseudomonas*)和葡萄球菌(*Staphylococcus*)中发现。它们的抗药性是使宿主对一些真菌所产生的抗生素产生抗性。通常这种抗药性自我遗传的。大多数R质粒是由相邻的二个DNA片段组成的(图17-15)。一个片段称为RTF(抗性转移因子, resistance transfer factor)。它带有调控DNA复制和拷贝数的基因及转移基因,有时还有抗四环素基因(*tet*)。其分子量为 $11 \times 10^6$ 。另一片段称为*r*因子或*r*决定子(*r*-determinant),这一部分大小不同(分子量从 $10^6$ 到大于 $100 \times 10^6$ ),它带有对其他抗生素有抗性的基因。常见的有抗青霉素(Pen)、氨苄青霉素(Amp)、氯霉素(Cam)、链霉素(Str)、卡那霉素(Kam)和磺胺(Sul)中的一种或几种。小的抗药性质粒缺乏转移能力,但仍带有抗四环素(*tet*)基因。如已

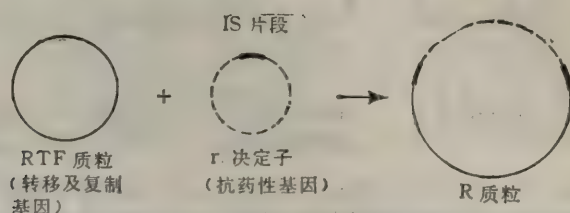


图 17-15 有感染性的R质粒的组成。在第十八章我们将介绍R质粒上的两个IS片段的形成机理:它们是由在*r*决定子中存在的一个单个的IS所形成的。



知的 pSC101 质粒,其分子量为  $5.8 \times 10^6$ ,常用于重组 DNA 技术(第二十章)。由两部分组成的 R 质粒就如同 F' 质粒。但抗性基因不是通过 RTF 的整合并接着异常切除而得到,而是通过转座子转移而来。这在第十九章将讨论。

小的抗性质粒,通常不称为 R 质粒。尽管它们十分小,但在革兰氏阳性菌中也有发现,如在金黄色葡萄球菌中发现的抗青霉素质粒。详细情况请参看本章的参考文献。

产大肠杆菌素的质粒或 Col 质粒

Col 质粒是能够产生大肠杆菌素的大肠杆菌质粒。它可以抑制不含 Col 质粒的菌株的生长。用字母来表示不同类型的大肠杆菌素(如大肠杆菌素 B)。每种大肠杆菌素都有抑制敏感细胞的特殊方式(表 17-2)。大肠杆菌素的检测方法与噬菌体检测方法十分相似。将一个产生大肠杆菌素的细胞放在一个铺有敏感菌的平板上。大肠杆菌素抑制周围细菌的生长,因而在混浊细菌层内产生一个清亮区——腔隙(lacuna)。

表 17-2 一些大肠杆菌的大肠杆菌素的性质

大肠杆菌素	大肠杆菌素的作用
大肠杆菌素 B } 大肠杆菌素 1b }	破坏细胞质膜
大肠杆菌素 E1 } 大肠杆菌素 K }	通过一种作用于膜上的未知效应 解除偶联的能量依赖过程
大肠杆菌素 E2	降解 DNA
大肠杆菌素 E3	分解 16 S rRNA

\* Col 质粒是细菌素生长质粒中的一种。这些质粒可在许多种细菌中产生细菌素。大肠杆菌素就是其中的一种。这些细菌素是蛋白质,结合在细胞壁上,抑制一种或几种基本的过程如复制,转录,翻译,或能量代谢。

大肠杆菌素的产生是可以诱导的,诱导因素就是那些使诱导性溶源菌产生噬菌体的因素,例如紫外线。如同诱导前噬菌体一样,产生大肠杆菌素的细胞也被杀死,但其机理尚不清楚。每个具有 Col 质粒的细胞群体中都有一小部分细胞也组成型地产生着大肠杆菌素。因此,Col 质粒的存在,对细胞群体的生存是有意义的。它可以使该细胞群体在与那些对大肠杆菌素敏感的细胞进行争夺共同所需的食物时处于更有利的地位。然而,这对于那些产生大肠杆菌素的单个的细胞来说,则是没有什么选择性意义的,因为产生大肠杆菌素后它们就将死去。

大肠杆菌素很可能是两种类型。一种是真正的大肠杆菌素,另一种,实际上是有缺陷的噬菌体颗粒。后一种是通过许多纯化了的的大肠杆菌素的研究而发现的。事实上,只有少数几种大肠杆菌素是简单的单体蛋白质。而其他的在电镜下观察则像是噬菌体的尾部。推测它们可能是从古代原始的前噬菌体的残留物上转录出的基因产物。有一种假说认为,可能由于多次重复地突变,结果噬菌体丢失了复制的基因、头部蛋白的基因以及裂解的基因,但是完整地保存下编码阻遏物的基因及尾部蛋白的基因。推测这些噬菌体具有 T4 的某些性质:它们吸附于细菌上,不需要将 DNA 注入,这种吸附作用抑制大分子物质的合成。

Col 质粒的分子量范围是几百万(不能自我遗传的质粒)一大于  $60 \times 10^6$  (可以自我遗传的质粒)\*。在 Col 质粒中研究得最多的是 Col E1,它的分子量是  $4.2 \times 10^6$ , 它被大量地应用于重组 DNA 研究和体外 DNA 复制系统中。

Col E1 是可以移动的,但是为非接合型的质粒。在遗传学研究中,它已用来证明从 Col E1 供体细胞到受体细胞需要一种质粒编码的核酸酶——其编码基因被称为 mob 基因,和一段特殊的碱基序列(移动的基础)。这个结论是从图 17-16 中所进行的实验而得到的。相容的质粒 F 和 Col E1 存在于一个细菌中。F 贡献出 Col E1 所没有的接合能力,所以它能使 Col E1 转移。(a) 中显示了在  $F^-$  细胞中 Col E1 的状态。mob 基因转录了, mob 基因的产物在 bom 位点产生缺刻,使 Col E1 由超螺旋环变为一个有缺刻的环。但由于 Col E1 没有产生性纤毛的能力,所以不能通过接合配对进行转移。又由于超螺旋状态与缺刻状态之间存在平衡。所以上述过程可以回复。在 (b) 中,细胞内有 F 质粒和 Col E1 质粒, F 可以合成出纤毛并产生转移器官。于是 Col E1 可以转移, Col E1 的 mob 突变体在 F 提供的转移器官存在下也不能转移。某些较大的质粒,例如 ColVK, 是带有产生大肠杆菌素基因的 F 质粒。遗传学分析发现 mob 编码一个蛋白质。进而分离突变体质粒,发现它没有形成松散的复合物。F 不能帮助  $mob^-$  突变体转移,见

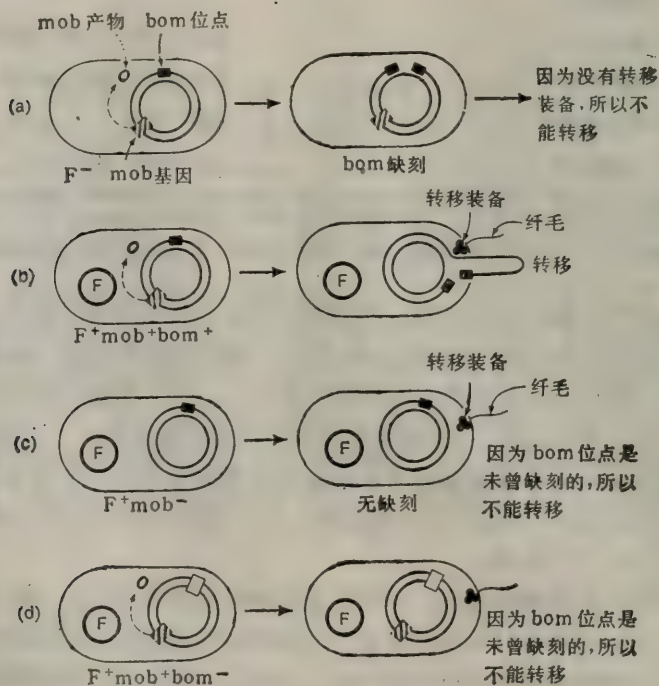


图 17-16 F 帮助 Col E1 质粒转移的条件。斜线区是 mob 基因及其产物。黑色的实心方块表示有功能的 bom 位点。空心框表示 bom 位点丢失或突变。只有在 ColE1 产生有活性的 mob 产物并作用于有功能的 bom 的位点, F 提供转移装备的情况下才可发生转移。

\* 某些较大的质粒,如 ColVK,实际上是携带着产生大肠杆菌素的基因的 F 质粒。



(c)所示。此时已形成 F 纤毛及转移器官,但是 Col E1 DNA 上不发生缺刻。(d)中又表示了另一个 Col E1 突变体,它有 Mob<sup>-</sup> 表型,是一个顺式显性突变。这种突变是一种缺失突变,丢失了 bom 位点, mob 蛋白质虽可形成,但不形成缺刻,它不能进行转移\*。

## 其他细菌质粒

还有两种质粒也需要特别提到。一些不同株的假单胞菌可将大量异乎寻常的有机化合物作为能量来源,因而细胞就需要极其大量的酶。对这些酶进行编码的要求已大大地超过染色体的编码容量。这些酶的许多基因存在于一些大的质粒上的。例如有一种能够利用樟脑为碳源的细菌所携带的质粒,就具有编码此种特殊用途的必要的酶的基因。不能做到这一点的细菌,并不是因为它是携带着缺陷型基因的突变体,而是因为它根本就没有这种质粒。自然界中没有单一的一种假单胞菌可以利用所有的碳源;不同的假单胞菌菌株,能对不同的物质进行代谢。分解不同物质的基因则是分布在形形色色不同的质粒上面。

在许多双子叶植物中发现了由根瘤土壤农杆菌所引起的冠瘿肿瘤。引起肿瘤的是一种称做 Ti 的质粒。在被感染的植物中,这种细菌进入植物细胞,在其中生长,并裂解,释放出它们的 DNA,此时细菌对肿瘤的形成已无作用。由于一种未知的机理, Ti 质粒中含有复制基因的一小部分整合进入植物细胞染色体中,这种整合的 Ti DNA 片段破坏正常时控制细胞分裂的植物激素的调控系统。结果正常的植物细胞变为肿瘤细胞。现在这种质粒已成为植物育种的一个重要工具。因为通过重组 DNA 技术可将特定基因插入 Ti 质粒,有时这些特定基因可整合进入植物染色体,从而永久性地改变植物的基因型和表现型。人们认为,可以用这种方式,把一些无亲缘关系的品种中的优良的、有经济价值的特性转移到已有的植物上,创造出植物的新变种。

## 真核生物中的质粒

在动物细胞中已发现环状 DNA 分子,并在实验室中保存于培养基中,它们可能是质粒,也可能是污染的病毒。这类因子的分布情况还知之甚少。可以预测在高等生物的细胞中,即使存在质粒,它们也将大大地不如在细菌中那样普遍。而且,在具有有性繁殖体系的高等生物的细胞中,几乎没有质粒 DNA 进行转移的机会。

在单细胞真核生物酵母中已发现质粒。杀伤微粒 (killer particle) 是最吸引人的一种。它是一个双链 RNA 分子,分子量为  $15 \times 10^6$ ,这是已知的唯一没有 DNA 的质粒。这个微粒有十个基因用于复制,还有其他一些基因用于编码合成类似于大肠杆菌素的杀伤性物质。

在酵母中研究最多是所谓  $2\mu\text{m}$  质粒。一个  $2\mu\text{m}$  长的 DNA 分子,其分子量为  $4 \times 10^6$ 。每个细胞中有 60 个拷贝,全部位于酵母细胞核内。这种位置表明,  $2\mu\text{m}$  质粒与酵

\* 编写本文时还没有得到关于 bom 位点就是转移起点以及 bom 突变不能形成有缺刻的松散复合物的证据,也没有得到关于转移是通过滚环复制而进行的证据;但已存在着单链分子在供体中重新环化以及这种环化分子可转移的可能性(尽管还没有强有力的证据,如 MB 噬菌体的释放)。在未证明它是不正确的之前,我们仍然必须考虑存在有这种可能性。

母染色体需要相同的有关复制酶,而且质粒 DNA 与核 DNA 一样也包裹有组蛋白。质粒 DNA 显然没有整合到宿主 DNA 中。

$2\mu\text{m}$  质粒的碱基序列包括 2 个反向重复序列,每个都有 599 个碱基对。它们被含有 2346 个和 2774 个碱基对的两个单一序列分开,如图 17-17 所示。在酵母细胞中有两种不同形式的  $2\mu\text{m}$  质粒,它们的区别在于其中一个单一序列的取向不同。这两种质粒的数量相同。如图所示,由于反向重复序列之间发生着频繁的重组作用,可以使质粒进行内部交换,质粒中促进这种内部交换的基因叫 *flp* 基因。关于这种内部转换作用的功能目前还不清楚。同时,在实验室中发现 *flp*<sup>-</sup> 质粒在宿主酵母细胞中的复制显然并没有缺陷。在自然界中内部转换作用是很有价值的,因为质粒的 40% 的编码能力都是在 *flp* 基因上。

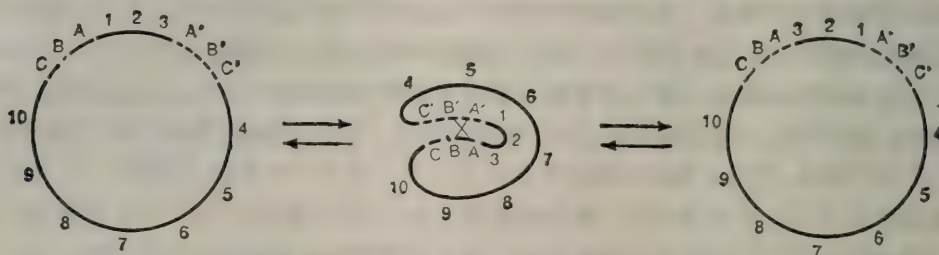


图 17-17 两种  $2\mu\text{m}$  质粒之间发生交换的机理。虚线片段是反向重复序列。四个片段的大小并不是按比例绘出的。

$2\mu\text{m}$  质粒在复制时有一个与其他质粒截然不同的特点,即它是一个高拷贝数的质粒。推测它的拷贝数受一个阻遏物体系调节,这个体系与前面我们介绍过的大肠杆菌质粒 pSC101 和 Col E1 有关。在前面谈到实行该种调控机理时,质粒是随机复制的。但是  $2\mu\text{m}$  质粒却不是那样,而是在宿主细胞的每一个分裂周期内,每个 DNA 分子都只复制一次,它是如何调控的目前还不清楚。

$2\mu\text{m}$  质粒为研究酵母 DNA 复制组了一个很有用的体系。在目前,它的最大用途是作为一个载体。通过重组 DNA 技术,它可以将外源基因克隆到酵母中去。这些将在第二十章讨论。

## 遗传工程中质粒的应用

第二十章将讨论 DNA 重组技术及其在遗传工程中的应用。这里只作简单介绍,着重强调质粒的作用。

限制性内切酶在质粒的非必需基因上产生一个单一切口,形成一个 DNA 片段。第二个线性片段甚至可能来自于另一种不同的有机体,它可以附着到上述被切割过的质粒的两个游离末端上,从而重新成环,环再被 DNA 连接酶焊接,形成封闭的 DNA 分子。我们设想:(1)加入的 DNA 是  $\beta$ -半乳糖苷酶基因的 DNA;(2)质粒的基因型是 *tet*<sup>+</sup>(抗四环素);(3)质粒有 50 个拷贝。这个外来 DNA 可以转化 *Lac*<sup>-</sup> *Tet*<sup>-</sup> 的细胞,于



是就可筛选到  $\text{Lac}^+\text{Tet}^+$  的细胞。如果质粒是高拷贝数的因子,这样,细胞中就有 50 个拷贝的  $\text{lac}$  基因,因此能产生大量的  $\beta$ -半乳糖苷酶。这个例子说明了质粒在基因工程中的一个重要用途——大量地扩增特定的蛋白质。

## 参 考 文 献

- Broach, J. R. 1982. "The yeast plasmid 2 $\mu$ m circle." In J. Strathern, E. Jones, and J. Broach, eds. *The Molecular Biology of the Yeast Saccharomyces*. Cold Spring Harbor Laboratory.
- Broda, P. 1979. *Plasmids*. W. H. Freeman and Co.
- Bukhari, A. I., J. A. Shapiro, and S. L. Adhya. 1978. *DNA Insertion Elements and Plasmids*. Cold Spring Harbor Laboratory.
- Clark, A. J., and G. J. Warren. 1979. "Conjugal transmission of plasmids." *Ann. Rev. Genetics*, 13, 99—125.
- Clewell, D. B. 1970. "Properties of a supercoiled DNA-protein relaxation complex and strand specificity of the relaxation event." *Biochemistry*, 9, 4428—4440.
- Clowes, R. C. 1972. "Molecular structure of bacterial plasmids." *Bact. Rev.*, 36, 361—404.
- Clowes, R. C. 1973. "The molecule of infectious drug resistance." *Scient. Amer.*, April, pp. 18—27.
- Deonier, R. C., and N. G. Davidson. 1976. "The sequence organization of the integrated F plasmid in two different Hfr strains of *E. coli*." *J. Mol. Biol.*, 107, 207—222.
- Gurney, D. G., and D. Helinski. 1975. "Relaxation complexes of plasmid DNA and protein. III. Association of protein with the 5' terminus of the broken DNA strand in the relaxed complex of plasmid Cole1." *J. Biol. Chem.*, 250, 8796—8803.
- Hayes, W. 1968. *The Genetics of Bacteria and Their Viruses*. Blackwell.
- Holloway, B. W. 1979. "Plasmids that mobilize bacterial chromosomes." *Plasmid*, 2, 1—19.
- Kingsman, A., and N. Willets. 1978. "The requirement for conjugal DNA synthesis in the donor strand during F'Lac transfer." *J. Mol. Biol.*, 122, 287—300.
- Kornberg, A. 1980. *DNA Synthesis*. W. H. Freeman and Co.
- Lewin, B. 1977. *Gene Expression. 3 Plasmids and Phage*. Wiley-Interscience.
- Low, K. B., and R. D. Porter. 1978. "Modes of gene transfer and recombination in bacteria." *Ann. Rev. Genetics*, 12, 249—287.
- Meynell, G. 1973. *Drug-Resistance Factors and Other Bacterial Plasmids*. MIT Press.
- Mitsubashi, S. (ed.). 1977. *R Factor: Drug Resistance Plasmid*. University Park Press.
- Novick, R. P. 1969. "Extrachromosomal inheritance in bacteria." *Bact. Rev.*, 33, 210—235.
- Novick, R. P. 1980. "Plasmids." *Scient. Amer.*, December, pp. 102—129.
- Willets, N., and R. Skurray. 1980. "The conjugation system of F-like plasmids." *Ann. Rev. Genetics*, 14, 41—76.

## 第十八章 遗传物质重排和交换机理 (II)

### 同源 DNA 序列间的重组

进化依赖于遗传物质在世代中的变异。遗传变异的第一步是突变。不同品系的突变作用在交配配对中结合到一起并在子代中表现为重组。经过许多世代过程,通过正常结合,使大量突变进行组合从而产生特异的变异品系。在双倍体生物中,这种重排发生在有性繁殖中。来自亲代的染色体在形成生殖细胞(卵及精子)时分离,在形成子代时又重新组合。

第二种变异也发生在形成生殖细胞的过程中,即配对的染色体拆开及重新组合。这个过程最初被称为交换(crossing over),现在称为重组(recombination)。在这一章和下一章将讨论重组发生的机理。

对于重组的具体机理,我们将采用研究微生物得到的成果进行讨论,即通过对细菌、噬菌体、酵母及真菌等微生物的研究所得到的结论以及其他的生化机理。虽然在细节上有所不同,但也都适用于高等生物。对酵母、真菌及高等生物遗传学有兴趣的读者可参看本章后的参考文献。

### 微生物中重组的类型

遗传变异在微生物进化中是非常重要的。由于微生物是单倍体,不同的等位基因通常不能存在于同一个细胞内。因此产生出一些可以使微生物从其他细胞获得遗传信息(即 DNA)的机理,或者说可以使外来 DNA,或 DNA 片段掺入到该微生物细胞的 DNA 中的机理。关于这个过程例子之一是转化作用(transformation)(第七章);死去的供体菌的 DNA 释放到周围环境中,然后进入受体菌。例子之二是转导作用(transduction)(第十六章),噬菌体通过生长在供体菌中而获取了供体细胞的 DNA,然后将此供体 DNA 注入到受体细胞中。最后的例子是接合作用(conjugation)(第十七章),Hfr 细胞 DNA(见第一章)或者通过细胞之间的接合与接触,从一个细胞转移到另一个细胞中。除了质粒的转移,凡进入到受体细胞中的遗传信息,必须掺入到新细胞的遗传器官中,才能在细胞分裂周期中成功地保存下来。在第十七章中,已知质粒是一个完整的复制单位,有自己的复制起点和复制循环,因而可以保存下来。质粒的保存下来是质粒 DNA 复制的结果。可是,在其他情况下,被转移的 DNA 不能继续复制。它们的保存必须通过插入到宿主 DNA 中,通常是通过取代机理而实现。这个插入过程一定需要受体 DNA 产生某种断裂然后使供体 DNA 与受体 DNA 重新连接,这个过程称之为遗传重组(genetic recombination)。

前面我们所介绍的前噬菌体整合(第十六章)就是一种遗传重组,即通过  $\lambda$  噬菌体 int 基因的参与发生位点专一性重组(site-specific recombination)。这个过程的特点是



在 4 个不定位点 (B, B', P 和 P') 旁的二个相同的短序列(15 个碱基对)间发生交换, 这种方式为专一性交换提供了一个重要的方法。

但是, 上面我们所提到的各种过程(转化、转导和通过接合而进行的 DNA 整合等)都要求较大的 DNA 片段进行交换(很可能为几百个碱基对), 而且它们的序列是相同的或接近于相同的。即两个发生作用的序列必须是同源的 (homologous)。因此, 这个过程也叫做同源重组 (homologous recombination)(图 18-1)。这类重组的另一个基本特点是必须有一定的宿主基因, 在大肠杆菌中这个基因就是 *recA* 基因。*recA* 基因的产物将在本章的后面部分具体地讨论, 它是 DNA 配对早期所需要的物质。在其他细菌中也存在类似的一些基因。所以这种重组类型称为依赖于 *recA* 型的重组。

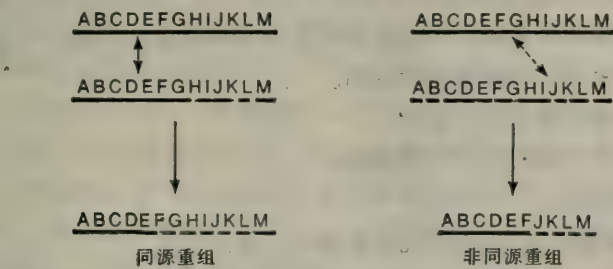


图 18-1 同源重组与非同源重组间的区别。横线为双链 DNA 片段。字母代表碱基对。虚线箭头指示发生遗传交换的位点。

重组的第三种类型发生在特定的 DNA 片段, 即转座因子 (transposable elements) 上。人们发现转座因子可以把一个 DNA 分子的一部分转移到另一个 DNA 分子上, 或者在双倍体细胞中把一部分 DNA 从一个染色体转移到另一个染色体上。这个现象称为转座作用。这是第三类重组——转座重组 (transposition recombination)。转座作用既不需要 DNA 片段有同源性, 也不需要 *recA* 基因, 因此, 转座作用是一种不依赖 *recA* 型的重组作用。

第四种重组作用, 既不需要 *recA* 基因的产物也不需要转座因子。它发生于非同源性序列之间, 被称为不正常重组 (illegitimate recombination)。我们对这种重组过程还完全不了解。

表 18-1 中列出了这四种类型重组的一些性质。

表 18-1 四种类型遗传重组的特性

重组类型	是否需要同源性	是否需要 <i>recA</i> 基因产物	是否需要序列专一性的酶
位点专一性重组	是	否	是
同源性重组	是	是	否
转座重组	否	否	是
不正常重组	否	否	?

本章只讨论同源性的或依赖于 *recA* 的重组, 第三及第四种类型的重组将在第十九章讨论。

## 同源重组的基本特征

在所有类似的重组作用中,重组的两个 DNA 分子或同一分子的两个片段发生相互作用,这种相互作用叫做配对 (pairing)。本章对同源重组的基本特征将讨论如下:

一些相应配对的蛋白质对于碱基序列本身并无专一性要求,不过它们确实需要具有相同的或接近于相同的序列的两个片段,以便通过配对在两个 DNA 分子中进行识别。

根据以上认识,用图 18-2 表示一个遗传重组的基本模型。在这个模型中,在不同链上的单链产生交错断裂,在互补链间形成碱基配对,使来自不同链的片段结合在一起。不幸的是,这种简单的模型不能解释遗传重组是如何发生的;但即使是更复杂的模型也无法解释许多普通的重组事件的特点。可是我们可以利用与检查其中的某一些模型,因为它们能说明人们是如何思考与理解重组过程的。

许多生物实验清楚地表明以下事实:

1. 亲代 DNA 分子产生断裂并重新连接。
2. 连接需要同源配对。
3. 重叠区(又称叠接区)在重组的几个阶段中出现。
4. 重组不需要 DNA 的复制,尽管在某些噬菌体系统中,只有通过重组后的复制才能进行包装;而且在高等生物中,类似于修复的复制作用对完成重组可能是需要的。

近年来,已建立了体外重组体系,而且这些体系已经帮助我们弄清遗传重组的一些步骤,如分支移动 (branch migration)、单链侵入 (single-strand invasion) 和反式切割 (trans-cutting)。进一步又发现了一些重组的缺陷型突变体,它们可以使体内重组在某一步骤上被阻断。所有这些结果都利于建立更精巧的、可测试的、符合实际的重组模型。对这些将简单地加以讨论。

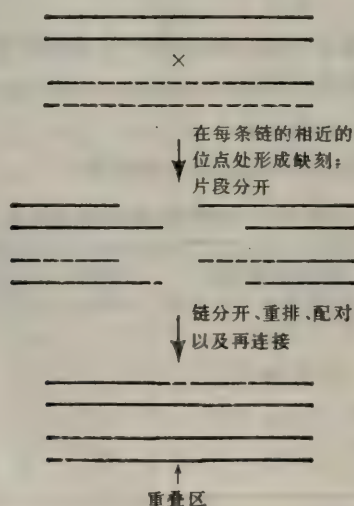


图 18-2 一个简单的但不很恰当的同源重组模型。实际上在重组子中存在着重叠区,但是在重组作用中相近位点形成缺刻并不是一个正常的阶段。



## 同源重组的特殊性质

在这一节我们将仔细讨论同源重组的几个特殊性质。在后面一节中通过对几个遗传现象的讨论,可知这几个特殊性质是目前流行的遗传学理论的组成部分。

### 断裂-重接及杂合现象

两个 DNA 片段在同源重组时,都产生断裂。通过碱基的互补配对,片段再重新连接起来。图 18-2 所示的就是一个有关断裂-重接现象的总体的、十分简单的(但又有启发性的)模型。要特别注意重叠区。在这个区域中,配对产生了重组分子,相互配对时并非每条链的序列完全一样。如果重叠区足够大,那么,并不要求在每条链中有相同的碱基序列,因为个别错配的碱基不会使整个结构的稳定性变弱,所以是可以忍受并修复的。的确,由于重叠区一般有几百到几千个碱基对,这种情况完全会发生。进一步,双链分子的单链发生断裂,产生出形成重叠区的两个部分,此断裂的位点也不必完全相同,如图 18-3。可以有如同(a)的情景:有缺口保留着,它可能有游离的  $3'-OH$  和  $5'-P$  基团。这些缺口可以被 DNA 聚合酶 I 填充并被 DNA 连接酶焊封。此外也可能出现(b)的情景,此时多余的 DNA 可被核酸外切酶切除。

利用这个极其简单的机理,在两个标记位点之间的重组情况如图 18-4(a)。如果这个交换包括三个遗传标记  $a^*$ ,  $b^*$  和  $c^*$ ,情况则如图 18-4(b) 所示。在这个例子中,由于断裂后  $b$  与  $b^*$  配对,使重叠区内有错配碱基对。这个重叠区是杂合的 (heterozygosis),

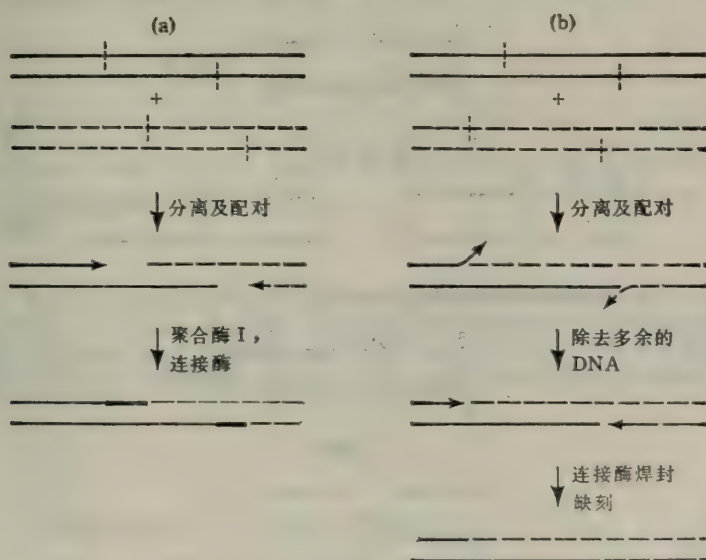


图 18-3 在不同位置切割单链的影响。用垂直虚线表示切割位点。(a) 连接遗留下缺口。这种缺口后再依次用聚合酶 I 和连接酶填补。(b) 连接遗留下多余的 DNA。箭头指出  $3'-OH$  基团。在每一行中,只指出了—个重组分子。注意,如果一个重组子具有缺口,那么其他的则一定具有多余的 DNA。

即  $b/b^*$  的杂合体(记住无论何时如有  $b/b^*$  杂合体存在,那么“外边”的标记都在一个重组重排体之内)。

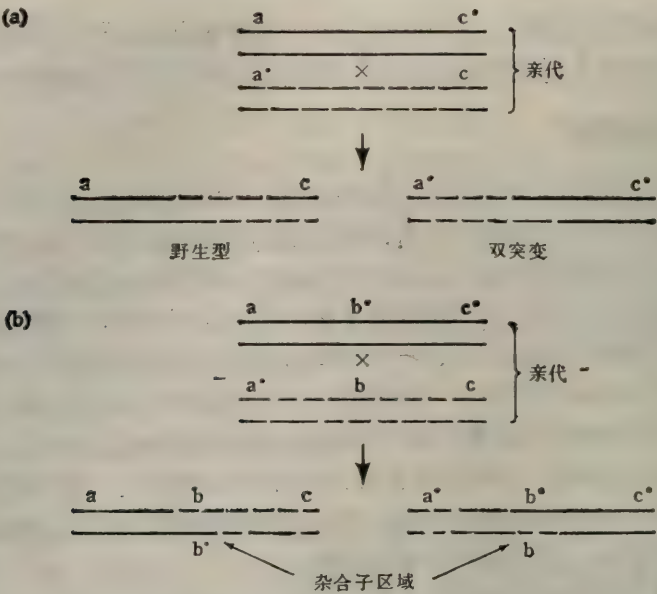
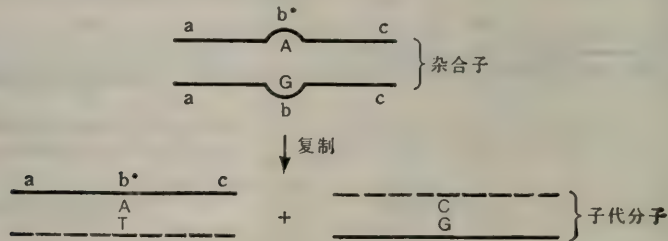


图 18-4 在带有遗传记号的亲代 DNA 之间进行重组的例子。(a) 在两个记号之间进行交换。(b) 跨越一个记号进行交换,因此产生杂合子后代。\*号表示等位突变。

(a) 通过复制作用使杂合子解体



(b) 通过错配修复作用使杂合子解体

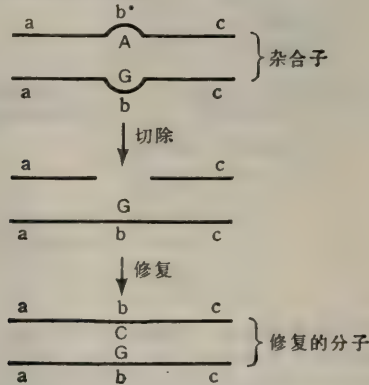


图 18-5 使杂合体解体的两种途径 (a) 通过复制。(b) 通过错配修复。



DNA 分子上的杂合性是不稳定状态。当基因型为  $ab/b^*c$  的杂合体进行复制时,  $b$  和  $b^*$  两个等位基因分别进入两个子代分子中[图 18-5(a)], 它们的基因型是  $abc$  和  $ab^*c$ 。杂合体状态也可通过错配修复(mismatch repair)被消除。即可以切除一般含有  $b-b^*$  错对中的那个碱基的序列, 再根据碱基配对原则加上一些碱基, 形成或是  $bb'$  对 或是  $b^*b'$  对, 见图 18-5(b)。图中符号用来表示正确的互补性。如果发生了这种情况, 那么究竟是形成基因型  $abc$ , 还是形成基因型  $ab^*c$ , 则取决于哪一条链被修复; 但是, 决不会同时存在以上两种基因型。这种错配的修复是极其普遍的。通常用子代在一个单一的遗传事件中失去一个遗传标记来观察这种修复。例如, 在上述所讨论过的交换中, 子代的基因型或者为  $abc$ , 这时  $b^*$  标记已丢失; 或者是  $ab^*c$ , 这时  $b$  标记被丢失。

### 分支移动

在第四章讨论过 DNA 的呼吸。呼吸是指 DNA 中的氢键随机地和暂时地打开与复合的现象。

分支移动(branch migration)是指碱基配对的双螺旋区中的一条链, 被另一条也可以配对的链所取代。这种现象的基础是上述呼吸作用。这可以通过图 18-6 中示出的检查人工合成的 DNA 多聚物实验来理解。分子中 W 链延伸, 链 I 和 II 可以与 W 链互补, C 链 I、II 结合后的长度超过 W 链。严格地说 C 链中的黑粗线区与粗虚线区具有相同的碱基序列, 它们都与 W 链的细虚线区互补。当发生呼吸作用时, 链 I 的右末端暂时打开, 可能发生二种情况: 链 I 回复或链 II 延长它的氢键区, 如 (b) 所示。当呼吸再次发生时, 可能 (a) 中分子回复, 这可能进一步产生变化如 (c)。呼吸作用从来不停止, 所以分子不同状态的变化也将继续下去。在后面将叙述的一个重组模型中, 分支移动起着重要的作用。

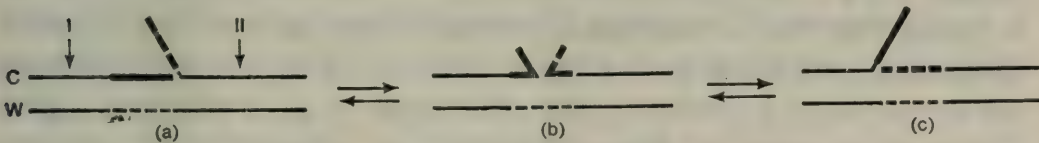


图 18-6 分枝移动。重区(粗黑线及粗虚线)都与 W 链中的虚线区互补。图 (a): 粗黑线区通过氢键束缚在虚线片段上。图 (b): 粗黑线区的游离端偶尔因两条链的张合(“呼吸”)而不受束缚, 这样就允许粗虚线区与细虚线区片段进行配对。图 (c): 通过连续的张合(“呼吸”), 更多的粗虚线区与虚线区片段进行配对, 或分子转变成图(a)的状态。当然, 也可能有许多许多像图(b)的中间状态类型。

### 单链侵入

单链侵入也依赖于呼吸作用, 设想一个环形 DNA 分子与一个单链 DNA 片段, 此片段与该环形分子有一段具有同源性的碱基序列, 见图 18-7。如果在同源区发生呼吸作用, 则此单链可能会与环形分子中的一段配对, 产生一个三链区。通过分支移动, 侵入的单链可能全被同化, 如图所示。在体外, 这样的单链侵入不会自动地发生。然而, 如果这个双链环状分子是超螺旋的, 那么与超螺旋相连系的解旋过程将增加单链泡(single-strand bubbles) 的大小并延长其存在时间(第四章)。但是, 如果温度升高至  $60-70^{\circ}\text{C}$  之

间，那么上述的单链同化过程将自动地、有效率地发生。细胞内的过程当然是不可能在如此高的温度下进行的；但是如果向研究体系中加入大肠杆菌 *recA* 基因产物，那么这种单链同化并进入超螺旋分子的过程就可以在常温下进行。我们很快就会看到，单链侵入可能是两个 DNA 分子进行配对的最早期步骤之一。催化这种相互作用的过程可能是 *RecA* 蛋白在同源重组中的重要功能之一。

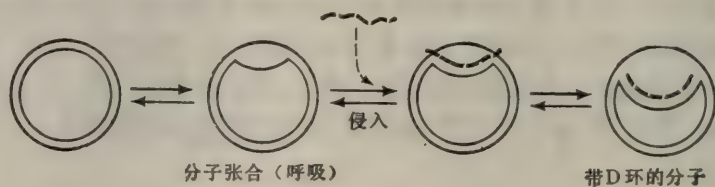


图 18-7 单链同化作用。互补单链侵入到一个环状 DNA 分子中并形成替代环(D 环)。

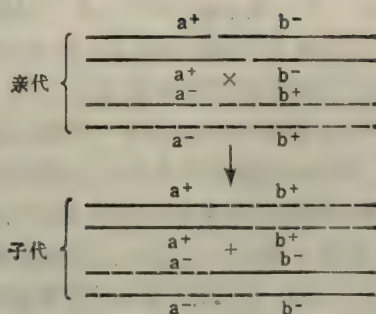
### 相互和非相互交换

考虑基因型为  $a^+b^-$  和  $a^-b^+$  的两个噬菌体之间的杂交。如果发生交换并回收得到  $a^+b^+$  和  $a^-b^-$  重组体，那么交换就是相互的 (reciprocal)，因为回收到相互重新配对的重组体。但如果只找到一种重组体，那么交换就是非相互的。遗传学的非相互交换和物理性的非相互交换之间，有着巨大的不同。图 18-8 可以说明这种不同之处。在 (a) 中，交换在  $a, b$  标记之间发生。这里没有丢失 DNA，交换是物理性的相互交换；人们找到了两种类型的重组体，因此这既是物理性的相互交换，也是遗传学的相互交换。在 (b) 中，在  $a$  标记的两侧发生断裂。第一排分子进行交换后的产物（见第二排）是物理性的相互交换分子。重叠区出现  $a^+/a^-$  杂合体。这是非常重要的，因为物理实验表明重叠区非常大。假如说没有错误配对，上述每个分子都复制，那就应当产生两种亲代基因型的分子及两种重组基因型的分子，这时的交换是物理性的相互交换。假如说错配修复发生在复制之前，并且在 (b) 的二种产物中，负的等位基因被转变成正常的等位基因 ( $a^- \rightarrow a^+$ )，最终的子代产物只出现一种亲代分子和一种重组体分子，如第三排所示。此时的交换为遗传学的非相互交换。实际上在许多情况下，物理性的相互交换可以通过多种方式表现出遗传学上的非相互交换；然而目前还没有一个简单模型可以使一个物理性的非相互交换产生出一个遗传学的相互交换。所以，物理性的相互交换可以是遗传学上的相互交换，或非相互交换；而物理性的非相互交换只可以是遗传学上的非相互交换，而不产生相互交换。因此，确定一个交换是否为遗传学上的相互交换，要注意防止机械的物理的观点，不能只以物理事件的现象来推出假说。

通常，重组作用是在很大数量的有机体上进行的，同时，这个过程往往掩盖了另一方面事实：一些单个个别的事件是非相互的交换。从图 18-8(b) 中看出，在最后一排中，错配修复产生遗传学非相互交换。从图中可推测，错配修复是使  $a^-$  变为  $a^+$ 。如果  $a^+$  变为  $a^-$ ，也可以看到非相互交换。假若错配修复随意地校正，就可能产生四种结果： $a^+b^+ + a^+b^-$  (非相互交换)， $a^-b^- + a^-b^+$  (非相互交换)， $a^+b^+ + a^-b^-$  (相互交换)，以及  $a^+b^- + a^-b^+$  (亲代型)；如果重组在很大数量的有机体里完成，并且如果对全部子代群体进行检测，那么  $a^+b^+$  与  $a^-b^-$  两种重组体的数目是相等的，表现为相互交换。事实上，当重组作用在群体上完成时，交换几乎总是表现为相互性的。非相互



(a) 物理性的和遗传学的相互交换



(b) 物理性的相互交换和遗传学的非相互交换

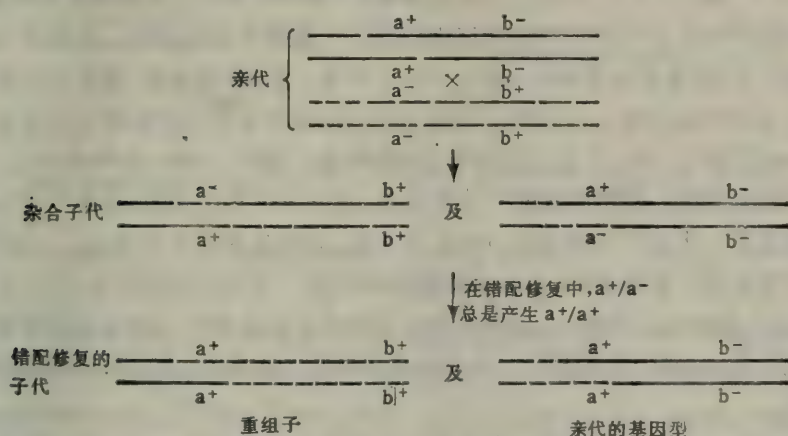


图 18-8 相互及非相互交换的含义。

交换性只是针对一些单个的重组事件的性质而言，而不是对重组事件总体而言。对于非相互交换，往往需要人们分离出单个的因素，才能检出。下面将介绍进行这种研究的两种方法。

## 单个重组事件子代的检测

酵母及其他一些真菌提供了一个检测子代分子发生某些类型重组事件的简单方法。这些生物的细胞，既有单倍体形式又有双倍体形式。当一个双倍体细胞进行减数分裂产生单倍体细胞时(酵母与多数真菌产生 4 个单倍体，少数特殊真菌产生 8 个单倍体)，同源染色体之间发生重组。对一种被人们普遍注意的子囊菌(ascmycetes) 生物进行研究发现，由一个双倍体细胞产生的 4 个单倍体细胞——子囊孢子(ascospores) 都存在于一个“袋”的子囊(ascus) 中；即每一个重组事件的产物分装在一个单独的子囊中，于是可以对单个重组事件进行分析。

下面是有关酵母的遗传过程。将已知基因型的双倍体细胞培养物涂布在一个无氮源的琼脂板上，这种过程将诱导产生减数分裂。在光学显微镜下观察，可以见到一些子囊。通过显微操作将这些子囊筛选出来并转移到第二个琼脂板上。于是子囊破裂，每个子囊

产生出4个子囊孢子,将子囊孢子一一进行机械性分离,任其分别长成一个一个克隆。再通过各种遗传学技术,可确定每一个克隆的基因型。这个冗长的过程称做四分体分析(tetrad analysis)。用这个方法可以获得单个个别的重组事件的全部产物。这显然是一种广泛适用于研究重组作用的最有效而又精确的遗传学技术。

在噬菌体系统中有一种与上面不同的称为单个爆破实验(single burst experiment)的技术。在这种实验中,大量的细菌被两种具有不同遗传标记的噬菌体感染。当宿主细胞被噬菌体吸附之后,将被感染细胞的培养物用培养基进行稀释,使每毫升中有0.05个被感染的细胞。然后按每个试管装1毫升上述悬液,将全部培养物分装于几百只试管中。依照泊松分布,在这个浓度下,95.1%的试管中没有被感染的细胞,4.8%的试管中有1个被感染的细胞,0.1%的试管中有1个以上的被感染细胞。因此,在含有受感染细胞的试管中,98%即4.8/4.9的试管中只具有1个受感染的细胞。继续培养使细胞裂解,将每个试管中的产物涂布于有指示菌的琼脂板上,于是将产生噬菌斑。接着是确定噬菌体的基因型是什么。如果发生重组(如果用于感染的是 $\lambda$ 及T4噬菌体,通常会发生重组),那么通过研究它们的基因型,可以得出感染期间所发生的遗传交换的数目及位置。一个典型的实验一般应研究100个受感染细胞,即意味着要使用2000个平皿,其中1900个平皿上将没有噬菌斑。这种实验除了浪费之外,另一个不足之处是,在单个爆破实验中受感染的细胞只有一半的子代噬菌体DNA被包装,某些基因型可能丢失了。而且,在一个已发生相互交换事件的细胞中,每次相互重组只产生1个拷贝的重组体,而一个重组体有50%的机会在爆破中不被释放,这有可能造成把相互交换判断为非相互交换的错觉。为了减少这个因素的影响,就需要研究更多的感染细胞的内含物,并且对这些实验数据进行统计学分析。

酵母和T4、 $\lambda$ 噬菌体都已通过上述方法广泛深入地研究过。这三种生物的非相互交换都是通过对数据进行统计学分析推算出来的。对重组机理的物理研究推测(但没有证实),DNA的物理的机械交换都是相互的,而非相互交换是由一些较后面的步骤,如错配修复等过程造成的。在下面的章节中我们将看到这些讯息所建立的几种重组模型。

## DNA 片段与染色体之间的重组

参与遗传重组的二个DNA分子可能是完整的染色体(环状的及线状的)、双链的片段或单链的片段。在这一节我们将讨论一个片段与一个完整的染色体之间的重组作用的某些性质。

### 细菌转化中的重组

在细菌转化中,染色体DNA与外源DNA发生重组。在肺炎球菌(*Pneumococcus*,已研究得最清楚)的转化中,当转化DNA进入细胞时,其一条链被分解,于是,只有一条链可与受体DNA重组。这条单链由于某种未知的原因可以不被核酸酶攻击——可能因它被单链DNA结合蛋白所包裹。在实验中用密度标记(重链)过的DNA去转化细菌(细菌的DNA是轻链),单链的供体DNA线性地插入到受体DNA中(图18-9(a))。通常被人们接受的关于这个过程的模型如图18-9(b)所示。在一种细菌蛋白质的帮助下(此蛋白



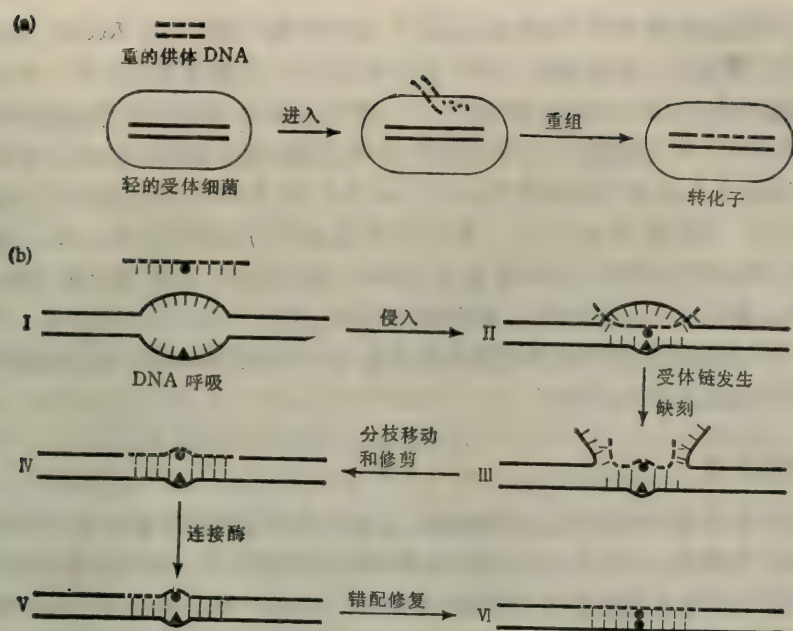


图 18-9 细菌的转化。(a) 双链供体 DNA 转变为单链，然后插入。(b) 单链供体 DNA 整合的一种假设机理。

很可能相当于大肠杆菌的 RecA 蛋白)，进入细胞的单链供体 DNA 片段插入到受体 DNA 中，由于一个未知的机理(参看下节反向切割 trans-cutting)，被取代的单链被切除。这个过程是非常有效的。然后发生分支移动，从而逐渐增加与受体链进行碱基配对的插入 DNA 的比例，修剪酶 (trimming enzyme) 再切除供体或受体 DNA 的游离末端。然后，利用连接酶使缺口焊封。最后，产生出一个带有错配碱基对的杂合区。假若此杂合区内有来自供体的标记，那么它能否在子代中出现，即能否在子代细胞中回收得到供体 DNA “重”标记，则取决于有无错配修复；如果有，那么取决于供体的“重”标记碱基或受体所得到的“重”标记碱基是否被切除。对于某些标记物，如高效标记物来说，或者错配总按照是供体的碱基序列被校正，或者很少甚至根本没有什么修复。在后一种情况下，细胞分裂之后，两个细胞中一个具有供体基因型，而另一个具有受体基因型。由于用于检测重组体所采用的培养基通常只允许重组后的细菌生长（如在一般培养基中加入抗菌素或使培养基缺乏某种必需氨基酸），所以，培养皿中产生的菌落，都是重组体细菌。对于低效标记物来说，错配修复通常从供体上移走错配碱基，使细胞仍为受体型。以上是关于两种类型标记物的正确解释。它的证据来自最近分离到的错配修复的、缺陷型的突变体。当这些突变体被转化时，所有标记物都像高效标记物那样地被转化，因为细胞分裂产生的两个子代细胞中总有一个细胞具有供体的基因型。

### 细菌接合中的重组

大肠杆菌 Hfr 和  $F^-$  细胞杂交时，单链 DNA 可以掺入到  $F^-$  的染色体中。重组体的形成需要 RecA 蛋白和一个二聚体的酶，它是 recB 和 recC 基因的产物，叫做 RecBC 蛋白。这个蛋白质有多种酶活性，其中包括一种很强的核酸酶的活性。图 18-10

所示的一个精巧的实验证明 RecA 蛋白先于 RecBC 蛋白起作用。RecBC 蛋白对于侵入链的插入是不需要的。如果具有  $lacZ^-$  基因型的 Hfr 细胞与携带着另一种  $lacZ^-$  突变的  $F^-$  细胞进行杂交,在杂交作用开始后不久,即可产生  $\beta$ -半乳糖苷酶。在没有形成  $lac^+$  重组体的  $RecBC^- F^-$  细胞中,也可以生成  $\beta$ -半乳糖苷酶。因此,认为上述供体与受体 DNA 必然是以互补双链的形式排列,  $lac^+$  mRNA 在 RecBC 蛋白起作用之前就已转录出来。如果  $F^-$  细胞是  $RecA^-$  的,那么  $\beta$ -半乳糖苷酶就不能合成。因此,链的侵入一定只发生在  $RecBC^-$  的细胞中,而不能在  $RecA^-$  的细胞中,并且 RecBC 酶一定是一种结束性的酶。到目前为止,RecBC 酶的功能仍然还是一个谜。

$Hfr \times F^-$  的重组过程的众多的特点预示着对于转化作用来说,其重组的机理并非像图 18-9 所示那么简单。

### 转导作用中的重组

双链 DNA 片段与染色体之间的重组,在特化的和通用的转导中都可以发生。其中转导的一部分线状双链 DNA 插入到细菌染色体上(见图 18-11)。注意,转导与细菌转化不同:转导中噬菌体在感染宿主细胞时把外来的双链 DNA 直接地注入细胞;而转化时,宿主细胞吸收外来 DNA 的过程必须将外来的双链 DNA 转变为内部的单链 DNA。

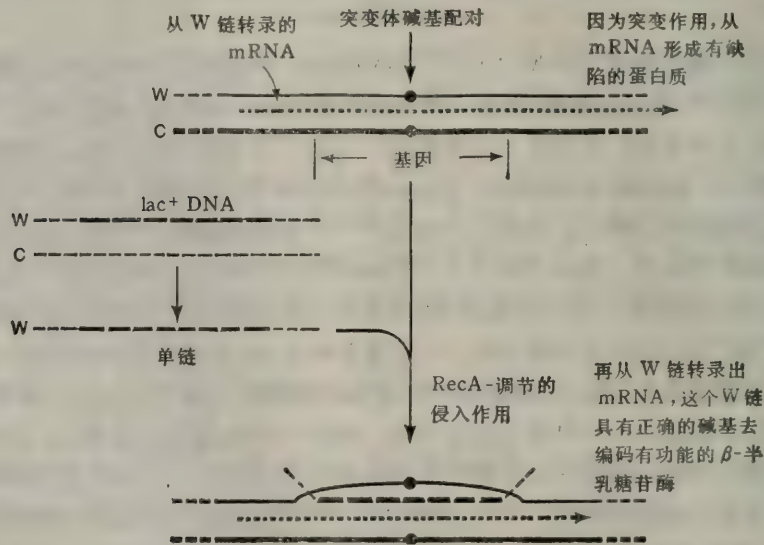


图 18-10 实验指出单链侵入需要 RecA 蛋白,不需要 RecBC 蛋白。长虚线代表 mRNA; 虚线上的粗线代表  $lacZ$  基因在编码链——W 链上的片段。



图 18-11 转导作用机理的略图。细节尚不清楚。



在大肠杆菌中进行转导时,对两种蛋白质: RecA 蛋白和 RecBC 蛋白都需要。这个过程与以前讨论的各种类型有显著的不同。 因为转导时不仅要发生二次交换,而且转 导的 DNA 的两条链都要 整合到染色体上,其 DNA 又是线状的,所以染色体双链中的两条链都要打开。有关转导作用中重组的机理还知之甚少,不能作更进一步的讨论。

### 反式切割

在图 18-9 建议的转化机理中,被取代的受体链上产生一个缺口对于受体 DNA 最终被供体 DNA 置换是必需的。我们将看到一些更普通的重组模型中这种缺口总是必然在某些阶段产生。近来却在被入侵的 DNA 上观察到发生缺刻的事实,这种现象称之为反式切割(trans-cutting)。在图 18-12(a) 的实验中,一个  $\lambda$  噬菌体溶源细胞被二个相同的  $\lambda$  噬菌体感染,它们与前噬菌体发生同源免疫作用(homoimmune)。由于复制作用被前噬菌体的免疫系统所抵制,进入细胞的噬菌体 DNA 径直地变成超螺旋状态。其中一个  $\lambda$  DNA 具有放射活性,另一个无放射活性但已预先用 psoralen 试剂处理。 当细胞被两个噬菌体感染后,用近紫外波段的紫外线照射,细胞中仅仅是经 psoralen 处理的 DNA 受到损伤。接着把生长在培养基中的受感染细胞经一个短暂的保温处理。人们发现,如果此细胞为 RecA<sup>+</sup> 并带有切除修复基因(参见第十章),那么其中带有放射性的 DNA 分子产生出缺刻。图 18-12(b) 是对这个实验的一种解释。经 psoralen 处理过的 DNA 因照射而造成的损伤(留下一段缺口)被切除。带有放射性的那个双链 DNA 分子在呼吸过程中(由于为超螺旋状态,所以它具有一个大的单链区,其解释在第四章),在 RecA 蛋白的促进下,被普素若仑(psoralen)处理过的 DNA 分子上的一段缺口中的单链侵入。为使被入侵的、带放射性的 DNA 分子中的单链区与入侵的单链的配对达到足够的稳定状态,核酸酶在其带放射性的另一条链上打出一缺口。尽管现在还没有分离得到这种核酸酶,但

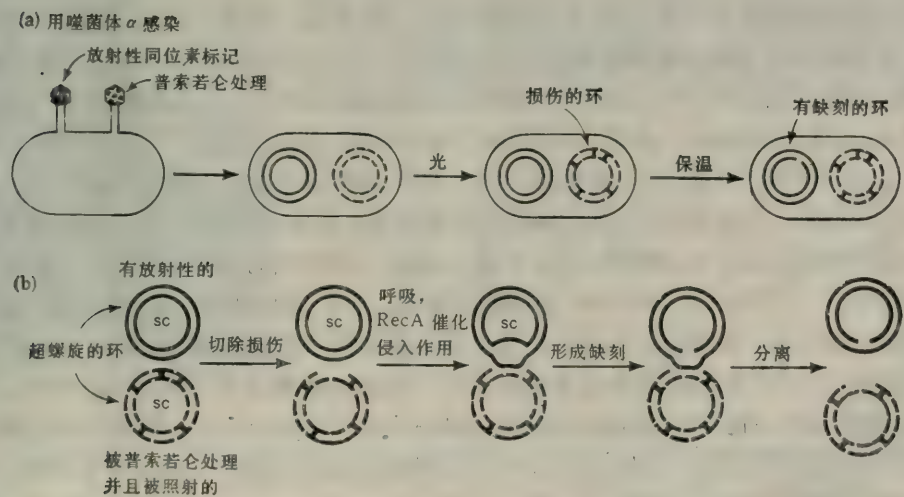


图 18-12 反式切割。(a) 证明反式切割的实验。(b) 对上述实验现象的解释说明。实线分子被虚线分子的单链侵入。以上解释还没有得到实验证明。超螺旋 DNA 分子在图中画成环状并标以 sc 字样。

目前已有很好的实验证据支持这种解释。

## 完整 DNA 分子之间的重组

在以上所讨论的例子中,参加重组的两个 DNA 分子中都有一个 DNA 片段。它们不如本节将讨论的完整分子之间的重组那么普遍和典型。除了几个例外,细胞中的完整的 DNA 几乎都是环状的,这就产生了一个我们还没有遇到的问题,即重组分子没有游离末端。目前,许多实验室都在研究这种类型的重组,并提出了大量的有关模型,我们将讨论其中的几种。现有两种不同模型:(1)在有一些环形 DNA 分子中首先打一个切口,切出一个游离末端,然后游离末端再侵入到未被打断的环状分子中,进行配对;(2)最近的新看法认为,同源配对的过程早于形成缺刻的过程。所有的模型都重视考虑以下事实:

1. 重组机理中包括一个重叠连接的步骤。
2. 重叠区非常大。
3. 重组作用既可以是相互的,也可是非相互的。
4. 错配修复是一个共同的、终结性的事件。除了不同的模式外,自然界中很可能至少存在着二种不同类型的同源重组。一种是依赖 RecA 蛋白类型。它发生在噬菌体、细菌及真菌体系中,不需要大量的 DNA 合成(指除了填补缺口 gap-filling 类型之外),利用 RecA 蛋白(在大肠杆菌中)或具有相似性质的其他蛋白。它可能是重组作用最普遍的一种类型。另一种是不依赖 RecA 蛋白类型,它依赖于复制作用利用一些编码噬菌体的基因,目前已在 $\lambda$ 噬菌体(Red 体系)及 T7、T4 噬菌体中对这种类型进行了仔细的研究,但还没有完全搞清楚。这一节只讨论依赖 RecA 蛋白的类型。

### Holliday 结构

有关重组的一个早期模型是 Robin Holliday 建立的。如图 18-13 所示,它假设存在着一个由两条单链进行重组而将两个分子连接到一起所形成的一个单位,这种结构被称为 Holliday 中间物(Holliday intermediate)(此处不讨论它是如何提出的)。如果在杂交(cross)区(Holliday 连接, Holliday junction)的对应位置上产生两个切口,那么配对的重组分子将产生如(c)的结果。另一方面,如果切点位于未破坏链的对应位点上将产生(d)的结果。(c)和(d)类型的分子都有重叠区,如果有遗传标记,那么这段重叠区可能会是杂合的。为了简单起见,将分子画成线状,如图中(a)到(d)部分。(e)部分表示配对的环状分子形成一个 Holliday 中间物的情况。“8”型结构是一个配对的环状分子所形成的 Holliday 中间物,这种分子已从具有多个拷贝质粒的细胞和受  $\phi$ X174 噬菌体感染的 RecA<sup>+</sup> 细胞中所分离得到的 DNA 通过电镜检查而观察到[图 18-4(a)]。进一步证实“8”形结构存在的证据,是用只在单一位点产生一个切口的限制性内切酶处理“8”型分子。如果“8”型分子中的二个环状分子是通过同源区连接的,那么处理后的分子应当转变为具有两对相同长度的臂的 X 型分子[图 18-14(b, c)]。为了进一步证实 Holliday 结构是重组中真正存在的中间物,要对“8”型连接的细节进行更详细的研究。

在图 18-13 的(b)行中所示的结构的一个环状组分可以按照图 18-15(a)所示那样



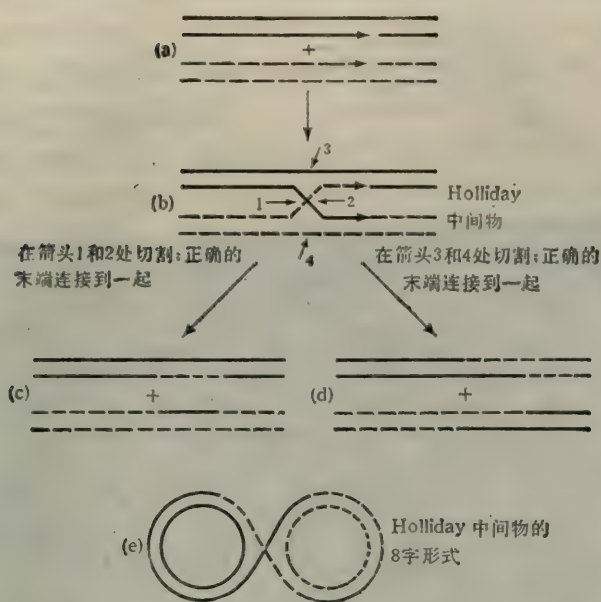


图 18-13 Holliday 中间物(b)的形成及重组分子的结构。两个带缺刻的 DNA 分子(a)转变为 Holliday 中间物(b), 形成两套重组子, 他们的结构见(c) 和 (d)。(e) 也是 Holliday 中间物, 与(b)一样, 只是它是环式的; 注意, 8 字型分子由两个单链的单体 DNA 环和一个单链的二体 DNA 环组成。

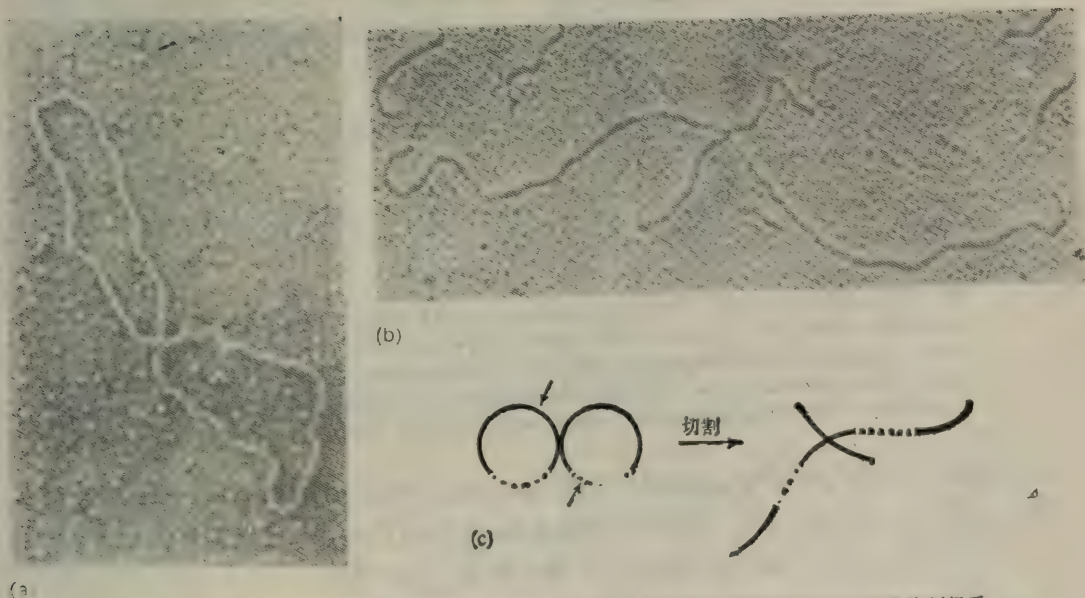


图 18-14 (a) 8 字型 DNA 分子 (8-shaped DNA molecule) 的电子显微镜照片。分子由两个  $\phi$ X-174 DNA 的单体环经重组作用形成。(b) X 型 DNA 分子 (X-shaped DNA molecule) 的电子显微镜照片。分子是由 8 型分子被限制性内切酶在特殊位置上切一下而形成的。(c) 是对 (a)、(b) 中分子的图解 (承 Robert Warner 提供)。

旋转  $360^\circ$ 。如果这个区域的单链组分都处于伸展状态, 那么结合点就会像图 18-15 中那样呈直角的样子。如果在甲酰胺存在的情况下 (使单链伸展) 通过电镜观察到了图

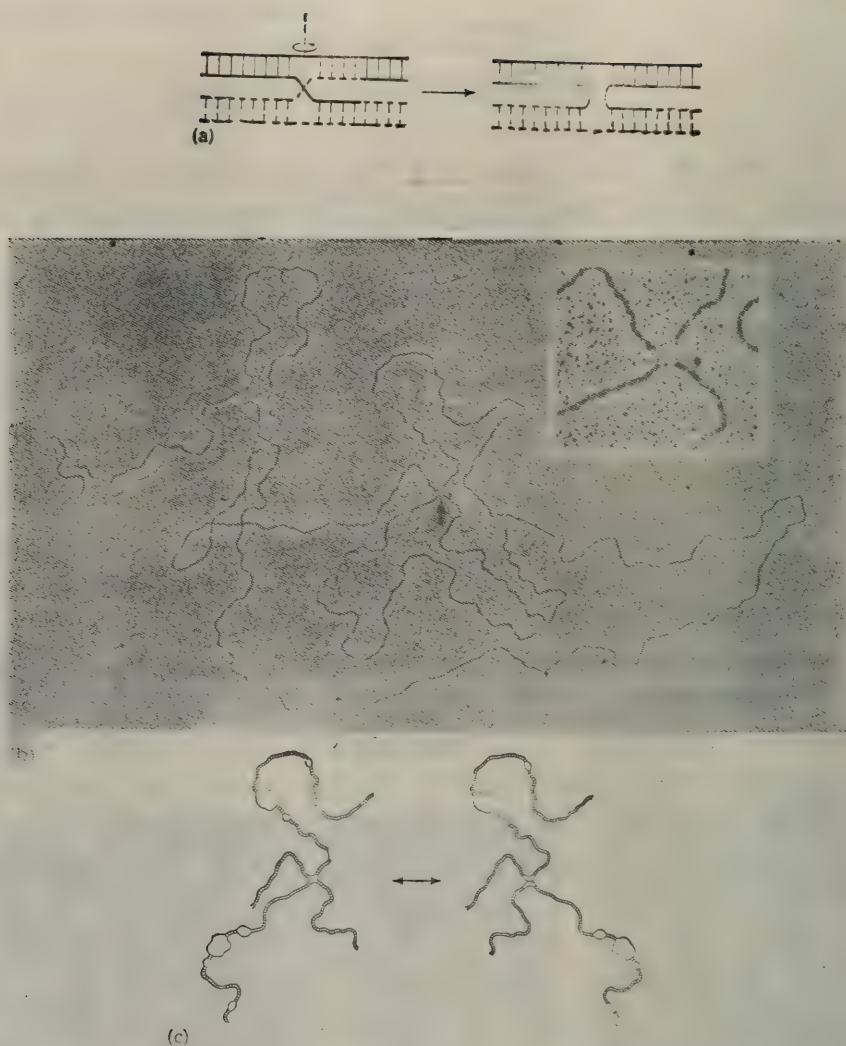


图 18-15 (a) 按图示的轴, 将 Holliday 结构的上半部进行旋转(握住下半部), 产生一个开盒式 (open-box) 结构。(b) 一个线性 DNA 片段与一个  $\lambda$  噬菌体的环状 DNA 之间发生重组作用, 产生中间物。图为该中间物的电镜照片。箭头表示 Holliday 结构的连接处, 右上角是这个连接处的放大照片。(c) 对 Holliday 结构周围区域的图解说明。左边的图表示开盒式构象(见 b); 右边的图表示一种可能的交换联会结构 (crossed structure)(承 Manuel Valenzuela 提供)。

18-13(b) 所示的 X 形结构, 也就找到了图 18-15(b) 中所示的那种分子。这些电镜照片再加上其他一些证据, 证实 Holliday 结构是 RecA 蛋白所介导的重组作用的中间物质。

下一节我们将对可能产生 Holliday 的一种途径进行讨论。

### 链转移或 Aviemore 模型

对于 Holliday 模型原来的观点认为链交换是一个断裂-转换-连接的过程, 在此过程中两条链在同一个位置处开始断裂。人们发现这是一个相互的过程。但是, 若干年后积累的事实证明交换并不总是相互的——最显著的发现是图 18-13 中见到的(c)行及(d)



行分别所示的来自两个亲代的充盈区 (filled-in region) 和重叠区 (overlap region), 通常仅仅是由相互作用的两个 DNA 分子中的一个分子衍生而来。因此, 人们花费了巨大的努力企图得到一个包含 Holliday 结构作为中介物的模型。其中的一个模型就是链转移模型, 我们就在下面加以介绍。

在解释链转移模型之前, 必须先明白图 18-16 中所示的 DNA 异构现象 (DNA isomerization)。异构化能使 Holliday 结构从 I 型变为 V 型。在解堆迭过程中, 每条双螺旋弯曲, 接近到 Holliday 接口处, 这样就使 Holliday 结构中的链都变成了直的。结构 II 中的右半边旋转  $180^\circ$  (由顶部转成底部) 形成结构 III。结构 III 的上面部分 (A 分枝与 b 分枝) 旋转  $180^\circ$  (由右端转成左端) 产生结构 IV。已弯曲的双链再堆积形成结构 V, 至此, 异构化作用已全部完成。注意, 在异构化的过程中链发生了什么变化——Holliday 结构中的交叉链变成了骨架链, 骨架链变成了交叉链。对于上面刚刚描述过的无论是广泛的解堆迭作用, 还是旋转性扩散作用, 都不能直接地在物理学实验中得到证明。然而, 采用分子模型则可以很容易地解释图 18-16 中所说明的弯曲与旋转, 而不会出现立体结构或张力限制等问题。

图 18-17 说明了链转移模型。其步骤如下:

1. 形成缺刻。通过一种未知的机理使双螺旋中的一条链断开, 缺刻处两个核苷酸分别带  $3'-OH$  基团与  $5'-P$  基团。在某种意义上讲, 细胞内 DNA 可能 (而且极为可能) 是不断地随机地形成缺刻 (被核酸酶作用) 并进行封闭 (被连接酶作用)。换句话说, 可能有一种机理使同源部分靠在一起, 而形成缺刻只为了对这种同源结合作出反应。注意这种模型只在一条链中形成缺刻, 而在 Holliday 模型中则要使两条链形成缺刻。

2. 取代。大肠杆菌 DNA 聚合酶 I 能作用于带游离  $3'-OH$  基因的缺刻, 向  $3'$  末端

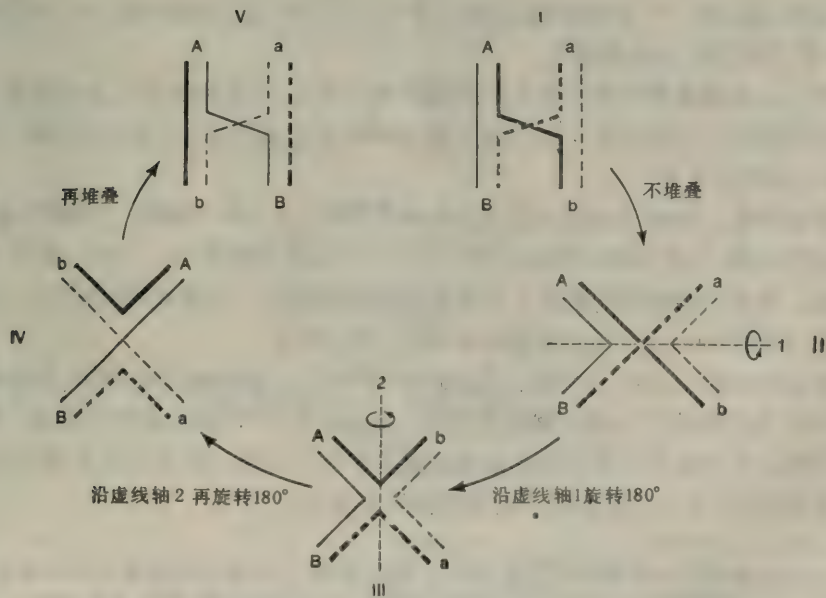


图 18-16 Holliday 中间物异构化的机理。I 型结构和 V 型结构可通过异构化互相转化 (与 M. Meselson 和 C. Redeling 1975 年在 *Proc. Nat. Acad. Sci.*, 72, 358 中提出的模型类似)。

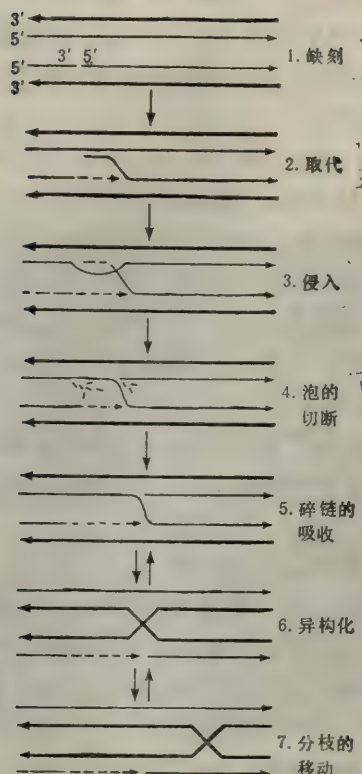


图 18-17 链交换成 Aviemore 模型。由 Meselson 及 C.Radding 提议(1975 年, *Proc. Nat. Acad. Sci.*, 72, 358)。虚线表示新合成的 DNA。

上加核苷酸,取代有 5' 末端的链(见第八章)。这种取代反应能无限进行,因而能形成很长的 5'-P 为末端的游离单链\*。

3. 侵入。游离单链可在双螺旋形成泡的地方侵入(泡是由张合反应或因有 DNA 结合蛋白而形成的)。很明显,这个过程需 RecA 蛋白的帮助。在以后的章节中将讨论 RecA 蛋白的作用方式。

4. 泡的切断。游离链的侵入使得被替换的那部分 DNA 降解。降解反应至少需要一种核酸内切酶,而且如果这种酶只有一个切点,那就还需要对 3'-OH 和 5'-P 的专一性外切酶。也有可能除去的是侵入链而不是被替换的链,它被外切酶除去。上述过程可能是反式-缺刻(trans-nicking)在起始步骤的一个例子。

5. 碎链的吸收。在第 4 步骤中形成的带 3'-OH 末端的缺口被 DNA 聚合酶 I 扩大,并由 DNA 连接酶将它与侵入链连接起来。同时,5'-P 末端的缺口可以进一步扩大,容许更多的侵入链与被侵入分子的其余部分碱基配对。注意,在两条双螺旋中仅仅只存在一个螺旋杂交区(含一条黑链和一条虚线链的双链区)。

\* 在这点上读者也许不清楚我们什么时候开始讨论了重组模型(这个题目是指不依赖于复制的重组过程)。由于三个原因,链转移模型被认为是能解释这种重组的模型。(1) 只发生少量 DNA 合成,实际上,只要延长 50 碱基对,对下面描述的插入过程就已是足够的了(虽然这个过程总共要涉及几百个碱基对);(2) 这种 DNA 合成不需要复制酶——DNA 聚合酶 III,并且可能只需要很少的复制因子;(3) 这种复制是修复复制。



6. 异构化。同图 18-15 中所示的异构化类型一样产生异构化作用。并且使结构中的几个臂端都放在重组构型中链交叉区的外翼位置上——形成 Holliday 结构。

7. 分枝的移动。分枝移动既可以向左进行也可以向右进行。在任何一种情况下，在两条双螺旋中都能形成杂交的双链区。

至此，已经得到了图 18-13(b)中所示的 Holliday 中间物，而且重组子还可形成图中后面几行所示的形式。重要的是 Holliday 结构中交叉链的交叉点可能距侵入的地方很远。所以，在图 18-17 的第 7 时期之后，发生缺刻的位置不一定非得在第 1 期的缺刻的所在位置上，即不一定如原初设想的 Holliday 模型所示的那样(图 18-13)。这就表明造成遗传交换过程的重组作用最终发生在两个遗传标记之间；然而，这种遗传交换的初始阶段的缺刻-侵入事件却不一定需要发生在两个遗传标记之间。

许多实验都支持链转移模型，但也偶然有一些现象与之并不一致，使得这个模型必需加以修改。

在下一节中将描述另一种完全不同的产生 Holliday 中间物的方法。

### 一个“先配对，后切割”的重组模型

链转移模型和其他许多模型都假定链切割过程发生在链配对过程之前。然而，有一个模型认为先发生链切割并无必要。在这个模型中(图 18-18)两个同源的区域先解链(或通过张合反应或在附加蛋白质的帮助下)，然后形成杂合的双螺旋。也就是说，如果在分子 1 中  $A_1 \cdot T_1$  是一对，在分子 2 中  $A_2$  和  $T_2$  是一对，杂合配对后就形成  $A_1 \cdot T_2$  和  $A_2 \cdot T_1$  的两种配对形式。其空间结构(图中 III)在主平面图中很难看清楚，但是此模型说明内部螺旋化的杂合双螺旋在结构上是可行的，并且可以靠氢链和碱基堆砌力得到稳定。在图

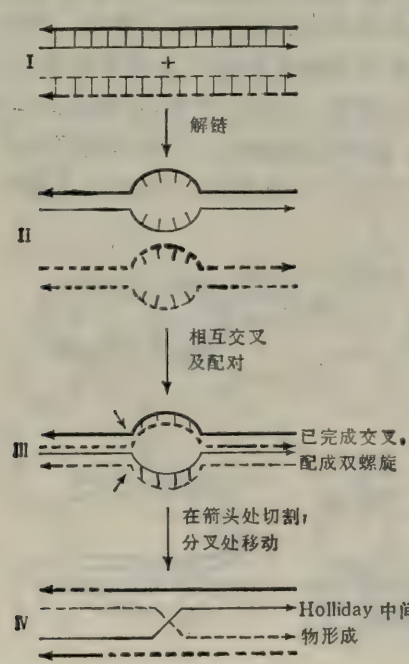


图 18-18 先配对后切割模型。

中内部螺旋上的箭头所指处打开两个缺刻可以产生 Holliday 中间物。这个模型有一点不利的地方,那就是这些缺刻必须产生在两条螺旋的相应位点上。一些最近的研究,如用几种拓扑酶如大肠杆菌 DNA 解旋酶 (gyrase) 和大肠杆菌解链酶 ( $\omega$  protein) (这些酶都有形成缺刻和闭合的活性,即有张合的功能)所做的实验表明,这些酶可能参与了这些过程。但只有很少的证据支持这个模型。很多实验说明带有缺刻的 DNA (或由外来试剂引起,或在修复碱基损伤的过程中形成)以及具有缺口或自由末端的 DNA 更易发生同源重组。这些结果都支持了链转移模型,并且不是简单地与“配对、切割”模型一致。然而,在没有得到相反的证据之前,我们还是要认真地考虑“配对、切割”模型。

8 型分子的解体

图 18-13 表明了由两条线状分子形成的 Holliday 中间物可以以两种不同形式解体。如果使双螺旋分开的切点位于二聚体中的链交叉区内,就产生类似于单链插入的形式(图 c);然而,如果切点位于别处,就产生具有重叠的连接部分的两个分子(图 d)。如果重组分子是环状分子二聚体而不是线状分子时,切割后也有两种可能的结果,但是这时产生的两种结果在结构上迥然不同。

图 18-19 给出了作用于 8 型分子的三种可能切割方式。为了清楚起见,我们把 8 型分子中的一个环旋转  $180^\circ$  形成一个新的结构。首先,我们假定这种结构很易断链而且只发生单链断裂。这意味着断链只能在图中四个箭头之中的一个所指的地方断裂。第二,我们假定断裂总是成对地发生,因为断裂处不足以使 8 型分子解体。这样就有三种再度配对的可能方式:(1) 在二聚体中链的相对部分;(2) 在单环中链的相对部分;(3) 在带链的环中的两个邻近点(一个在单环上,另一个在两聚体的链上)。这样三种再度配对方式可产生三种结果:两个单体,一个二聚体,一个滚动环,如图 18-19 所示。这三种不同形式可能在不同的系统中有着很不同的功能。例如,如果重组事件的最原本的接受者是只有单体长度的真核染色体或细菌染色体,就既不能形成二聚体也不能形成滚动环,也不能进行其他必要的进一步进程。而另一方面,在噬菌体系统中,如果重组过程只形成单体,那么将最终导致死亡。例如  $\lambda$  噬菌体,由于包装需要两个 *cos* 位点(第十五章),所以只有形成二聚体或滚动环(可以不断地复制)才能形成重组的  $\lambda$  噬菌体。但是目前还没有

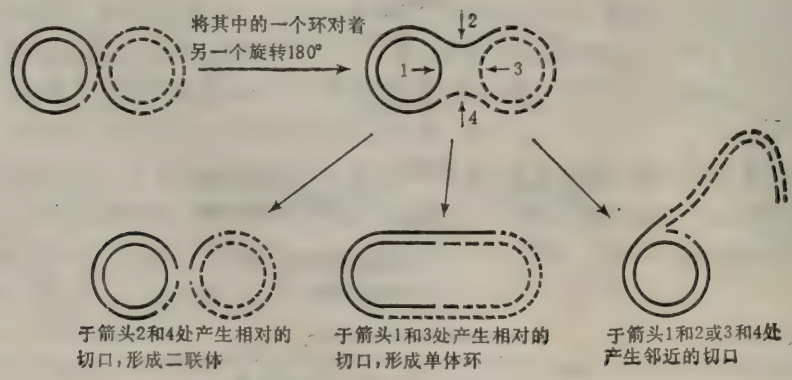


图 18-19 8 型分子的解体。



在任何系统中发现任何有重组作用介入的滚动环的证据。然而，在 *recA*<sup>+</sup> 细菌中经常可以发现小质粒的二聚作用。毫无疑问，8 型分子至少在遗传重组中是一种中介物，但是现在对于它是如何形成重组分子的以及它是如何转变成重组分子的则还不清楚。

## RecA 和 RecBC 蛋白的性质

在大肠杆菌中存在着几种同源重组的途径。最主要的一种途径需要有 RecA 和 RecBC 蛋白以及可能还需要其他一些蛋白，如单链 DNA 结合蛋白（这个蛋白质还参与其他反应）参与。人们已经能得到 RecA 蛋白及 RecBC 蛋白的纯品，而且花了很大力气对它们的 DNA 结合活性和酶活性进行研究。这些研究的结果已暗示出这些蛋白质的生理功能，但还没能给出很清晰的讯息。

### RecA 蛋白

RecA 蛋白有两个主要的化学活性：（1）它是一个依赖于 ATP 的单链结合蛋白。（2）它是一种蛋白水解酶。其蛋白酶活性在调节它自身的生物合成中是必须的，因为它可以水解 LexA 阻遏物（看第十四章），所以它的 DNA 结合活性与重组作用有关。

RecA 蛋白的主要功能是催化游离单链的侵入，图 18-20 表明了好几种在体外进行的、由 RecA 蛋白介导的 DNA 与 DNA 的相互作用。图中所示的各种结构是稳定的，由同源序列之间的正常的碱基配对而靠拢到一起。证明这个结论的证据，就是这些结构只有在被加热到 DNA 的正常融解温度时才能解离。图中的每种反应都需要两种 DNA 分子、RecA 蛋白（不需其他蛋白质）和 ATP。在反应中 ATP 被水解，这意味着链的侵入是一种耗能过程，但其中 ATP 的具体作用还不清楚。

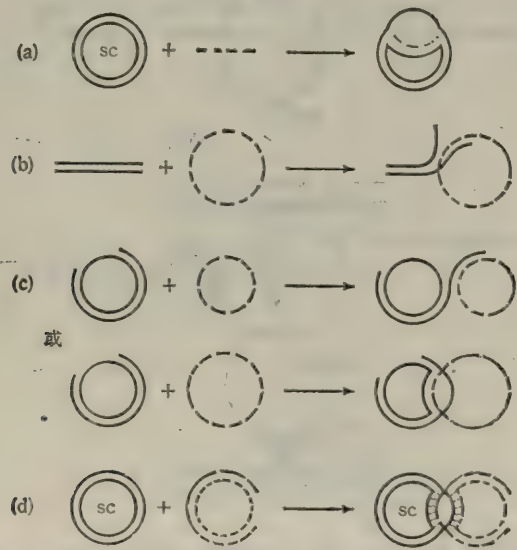


图 18-20 RecA 蛋白调节的四种反应。sc 表示超螺旋。

图中所示的四个反应有三个要求:

1. 其中一个分子必须是单链分子,或者有足够长度的单链区,这种分子用虚线表示。
2. 另一个分子(用实线表示)必须含有游离末端或者单链区。在图(a)和图(d)中,共价环的单链区由超螺旋提供,如果不是超螺旋分子,a、d中的相互作用就不会发生。
3. 其中一个分子必须有自由的(游离的)末端。

这种相互作用的详细机理还不清楚,但可以简单地认为 RecA 蛋白能使一个单链分子的自由末端侵入到另一个双链 DNA 分子中的同源单链区。由于 RecA 蛋白没有解旋活性,它不能使被侵入的分子中产生单链区,这样就解释了为什么需有第二个要求。然而,更精细的实验(图 18-21)表明,在两个 DNA 分子之间的原初的相互作用并不需要第二个要求,这与图 18-20 中所示的为了形成该图中稳定的结构而需要一个自由末端的要求相反。

这个实验利用了两种 DNA 分子:(1) G4 噬菌体的双链复制型 DNA (约 6 000 碱基对),用限制性内切酶把它切成线状双链 DNA 分子;(2) 由插入的 274 碱基对的 G4 DNA 序列插到几个千碱基对的 M13 的 DNA 序列中,组成 M13-G4 单链的杂交体分子(见图 18-21 上部)。这两种分子只有 274bp 碱基对的 G4 序列是同源的,所以 G4DNA 分子的双链末端则和 M13-G4 杂交分子的任何地方都不可能发生相互作用。然而,RecA 蛋白却能使这两种分子发生相互作用。这说明对于两个分子的靠拢与联结来说,并不需要有自由的同源末端。这个结论又被下面的一个实验证明。在这个实验中,使用了未切开的超螺旋的 G4DNA 分子,G4DNA 与 M13DNA-G4DNA 靠拢在一起。M13-G4 与 G4 的复合体与图 18-12 中所示的那些复合体有一个重要区别,那就是 M13-G4 与 G4 形成的复合体不很稳定,在远远低于 DNA 熔点温度(20℃)的情况下就解离了。这个结果表明这种相互作用与图 18-20 中所示的不同。显然,M13-G4

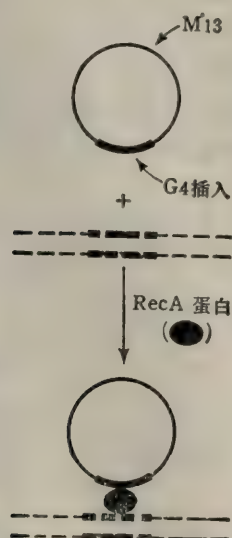


图 18-21 大分子 DNA 通过 RecA 蛋白介导与线性 G4DNA 的结合。这种结合是借助于携带有 274 碱基对的 G4DNA 插入序列的单链 M13 而完成的。M13 的其他部分与大分子 DNA 都是不同源的,只有粗线的部分(G4DNA)能够碱基配对。



与 G4 的相互作用不能形成大段的互补区，因为那样形成的 DNA 必须解链，不能形成螺旋（记住：螺旋的每一圈只有 10 个碱基对被拆开）。然而，相互识别又一定要借助于碱基配对。上述反应形成的产物的热稳定性较差，也表明碱基互补区非常短，多半短于 10 个碱基对（密码子与反密码子的相互作用只需要三个碱基对）。所以这三条链之间的相互作用也可能符合图 18-18 中图 III 所示的“先配对再切除”的模型，但在完全弄清相互作用的实质之前还需进一步深入研究。

上述实验的重要性就在于它明确指出了 RecA 蛋白能够介导不需 DNA 的自由末端的较弱的相互作用。因此说明，链与链的识别并非必需有自由末端。推测，相互作用的区域可沿着双螺旋移动（通过某种分枝移动），直到遇上一个自由末端；然后，从自由末端处，发生解链，如图 18-20 所示，接着在由各个反应分子所提供的一条链与一条链之间就会形成具有许多碱基配对的 Watson-Crick 双螺旋。人们可以进一步去推测已在许多生物系统中人工形成的单链缺口会引起刺激作用，它可能会缩短弱相互作用的 DNA 分子在解离之前遇见自由末端并形成稳定的双螺旋所需要的时间。这种刺激作用已经在许多种生物体系中有记载。

至此，我们还没有谈到 RecA 蛋白怎样参与完成所有这些过程的。原因很简单，那就是虽然最近的发现给我们不少启示，但我们还并不知道 RecA 蛋白究竟是怎样完成这些过程的。已经发现在一定的条件下，单链 DNA 分子、双链 DNA 分子和 RecA 蛋白之间也会形成弱的复合体，甚至当两种 DNA 分子之间没有同源性时也会形成这种复合体。这暗示出，当两种 DNA 分子之间有同源性的时候，RecA 蛋白使两个 DNA 分子之间的任何区域发生相互作用，然后允许这些分子作前后移动，直到遇见同源区为止。

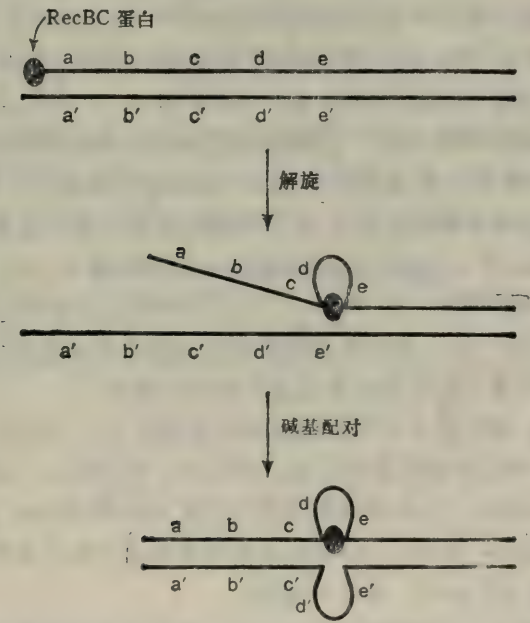


图 18-22 RecBC 蛋白质帮助双链 DNA 解链的模型。

## RecBC 蛋白

RecBC 蛋白是一种多功能的蛋白质。它具有内切酶活性以及很强的作用于单、双链 DNA 的外切酶活性。由于首先发现它具有外切酶活性, 所以这种蛋白又称作核酸外切酶 V, 在一定的实验条件下可以抑制其核酸酶活性, 这时这种酶就表现出使双链 DNA 解旋的能力, 解旋的同时发生 ATP 水解。当这种核酸酶活力被抑制的时候, 使 DNA 发生解旋的现象是不寻常的, 像图 18-22 所示那样。酶与 DNA 分子的一端结合, 使 DNA 解旋。它沿着 DNA 的一条链移动, 但与这条链的两个区域保持接触, 使这条链的末端穿过酶, 这样就形成了如图(b)所示的单链环。当单链尾巴的长度不断增加时, 由于与另一条链之间发生同源互补, 就形成了如(c)所示的双泡结构。这可能就是体内产生单链区的一种机理, 而不可能是以下的那种为 *recA* 基因所介导的配对作用提供单链区域的机理。因为前面的实验看来暗示出 *RecA* 蛋白的作用已先于 *RecBC* 蛋白的作用。如果核酸酶活性未被抑制, 那么在单链上相距几千个核苷酸之处会发生缺刻。*RecBC* 蛋白也能与 DNA 聚合酶 I 形成复合体, 但此复合体的性质还不清楚。*RecBC* 酶的酶活性已证明在许多同源重组的模型中有作用, 但还没有实验对它在重组作用的特殊阶段中的精确作用作出证明。

## 缺乏重组蛋白体系的性质

缺乏重组能力的突变细胞有一种另外的现象, 即他们对自己的 DNA 损伤过度敏感。例如, *recA<sup>-</sup>* 或 *recBC<sup>-</sup>* 的细菌都可能被大大低于正常细胞(*rec<sup>+</sup>* 细胞)致死量的紫外线杀死。有关 *recA<sup>-</sup>* 突变的数据可在第九章的图 9-10 中看到。敏感的主要原因是, 在 *recA<sup>-</sup>* 或 *recBC<sup>-</sup>* 细胞中缺乏一种主要的修复途径——重组修复。

*recA<sup>-</sup>* 细胞的 DNA 是不稳定的——它们不断地被降解而又不断地重新合成, 用紫外线照射后的最显著的特点就是 DNA 发生降解; 然后, 一半以上的细菌 DNA 再被 *RecBC* 酶促降解成短的核苷酸片段。这种猛烈的降解称作粗率 (reckless) 现象。

*recA<sup>-</sup>* 细胞的另一个特征就是它们比 *recA<sup>+</sup>* 细胞生长得慢, 这表明它还有其他缺陷, 但人们还不了解它的大部分缺陷。*rec<sup>+</sup>* 大肠杆菌在丰富的培养基中, 于 37°C 下大约 25 分钟就分裂一次。相比之下, 典型的 *recA<sup>-</sup>* 突变体细菌则需 40—60 分钟。生长慢的一个原因可能是由于粗率降解现象。而 *recBC<sup>-</sup>* 在生长中表现的缺陷更是极端严重, 它需要 100—200 分钟才能分裂一次。这些生长缺陷无疑地是由许多过程造成的, 即 *rec<sup>-</sup>* 细胞中有许多过程或者是根本不起作用或者是功能性的效率低下。

在噬菌体体系中, 重组的能力有时与 DNA 复制有关。例如, 沙门氏杆菌噬菌体 P22 含有一个线性 DNA, 但在其复制之前必须环化。P22DNA 分子不含有互补的单链末端, 但它是末端冗余的。冗余末端序列之间的重组使得 DNA 分子环化。这个反应由 P22 的 *erf* 基因催化, 而 *erf<sup>-</sup>* 的琥珀突变噬菌体又不能生长在那些缺乏琥珀校正子 (amber suppressor) 的 *recA<sup>-</sup>* 宿主细胞中。

大肠杆菌  $\lambda$  噬菌体是一个典型的例子, 它的一些重组基因起着生命攸关的作用, 虽然还不完全清楚这些基因的主要功能是否就是关系到重组作用。例如  $\lambda red<sup>-</sup>$  噬菌体, 它在



*recA*<sup>-</sup> 宿主细胞中不能产生重组的子代噬菌体。 $\lambda$ *red*<sup>-</sup> 噬菌体虽然仍然能够生长,但是,在受它感染的 *recA*<sup>+</sup> 宿主细胞或 *recA*<sup>-</sup> 宿主细胞中, *red*<sup>-</sup> 噬菌体产生子代噬菌体的数量只是那些受 *red*<sup>+</sup> 噬菌体感染的细胞中产生的子代噬菌体数量的一半。因此重组作用(或者至少是  $\lambda$  的 *red* 基因的功能)对于产生大约 50% 子代噬菌体是必需的。 $\lambda$  的另一个叫做 *gam* 的基因的作用主要是在感染早期使宿主的 RecBC 蛋白失活。*red*<sup>-</sup>*gam*<sup>-</sup> 噬菌体比 *red*<sup>+</sup> *gam*<sup>-</sup> 噬菌体的生长能力更差。如果宿主是 RecA<sup>-</sup> 型的细菌,那么 *red*<sup>-</sup>*gam*<sup>-</sup> 噬菌体感染它后根本就不能产生子代。但是宿主同时又是 RecBC<sup>-</sup> 的时候,产生子代的能力就能得到恢复,这就说明有活性的 RecBC 蛋白的功能是破坏  $\lambda$  的发育。因此,当有功能的 *red* 基因的产物与无活性的 RecBC 蛋白同时存在时,  $\lambda$  能生长得更好些。我们并不想阐明这个非常复杂的现象(已有许多模型,但都不很适当),只是要指出: 这些突变作用中的每一种(*red*<sup>-</sup>、*gam*<sup>-</sup> 以及 *recA*<sup>-</sup>)都以某些方式使 DNA 合成发生失常。例如,当用  $\lambda$ *red*<sup>-</sup> 突变体去感染大肠杆菌时,在感染周期完成之前, $\lambda$  的 DNA 复制就被关闭。当用  $\lambda$ *red*<sup>-</sup>*gam*<sup>-</sup> 去感染 *recA*<sup>-</sup> 宿主细胞时, $\lambda$  噬菌体的 DNA 滚动环复制受到严重损害,同时也发现有一些不正常的复制产品。关于这些复制产品的结构细节以及缺少的那些分子的细节都还不清楚,但有一点是清楚的,那就是除了对遗传记号进行重组之外, $\lambda$  中的这些重组基因还有其他的功能。

## 大肠杆菌的其他重组体系

前面我们已经提到,大肠杆菌有几种同源重组的途径,每一种途径都需要 RecA 蛋白,刚才讨论的那种还需要 RecBC 蛋白。当用 Hfr 的细胞与 *recBC*<sup>-</sup> 型的 F<sup>-</sup> 的细胞进行杂交时,可以证明还存在有其他的重组途径。如果 F<sup>-</sup> 细胞是 *recA*<sup>-</sup>, 重组频率就下降 10<sup>6</sup> 倍,而如果 F<sup>-</sup> 细胞是 *recA*<sup>+</sup> *recBC*<sup>-</sup>, 重组率只下降 100 倍。在 *recBC*<sup>-</sup> 的细胞中发生的重组依赖于几个基因;研究得最清楚也最重要的是 *recE* 与 *recF* 基因(下面的讨论很复杂,需要经常地参考表 18-2 和图 18-23,初学者则可跳过本段看下一章)。在正常细胞中 RecF 途径效率极低,基本上被核酸外切酶 I 抑制,外切酶 I 是由 *sbc* 基因合成的图 18-23 说明了这个现象,据认为其过程如下: 在通过 RecBC 途径形成重组子时,一个重要的中间物“Y”被 RecBC 蛋白变成重组子。这个转变也许可以分为几个步骤。可以

表 18-2 几种大肠杆菌突变体的重组能手

基 因 组	是重组能手吗?	活性体系
<i>recA</i> <sup>-</sup>	不是	没有
<i>recBC</i> <sup>-</sup>	不是	没有
<i>sbcB</i> <sup>-</sup>	是	RecBC
<i>recBC</i> <sup>-</sup> <i>sbcB</i> <sup>-</sup>	是	RecF
<i>recBC</i> <sup>-</sup> <i>recE</i> <sup>-</sup>	不是	没有
<i>recBC</i> <sup>-</sup> <i>sbcA</i> <sup>-</sup>	是	RecE, RecF
<i>recBC</i> <sup>-</sup> <i>sbcA</i> <sup>-</sup> <i>recE</i> <sup>-</sup>	不是	没有

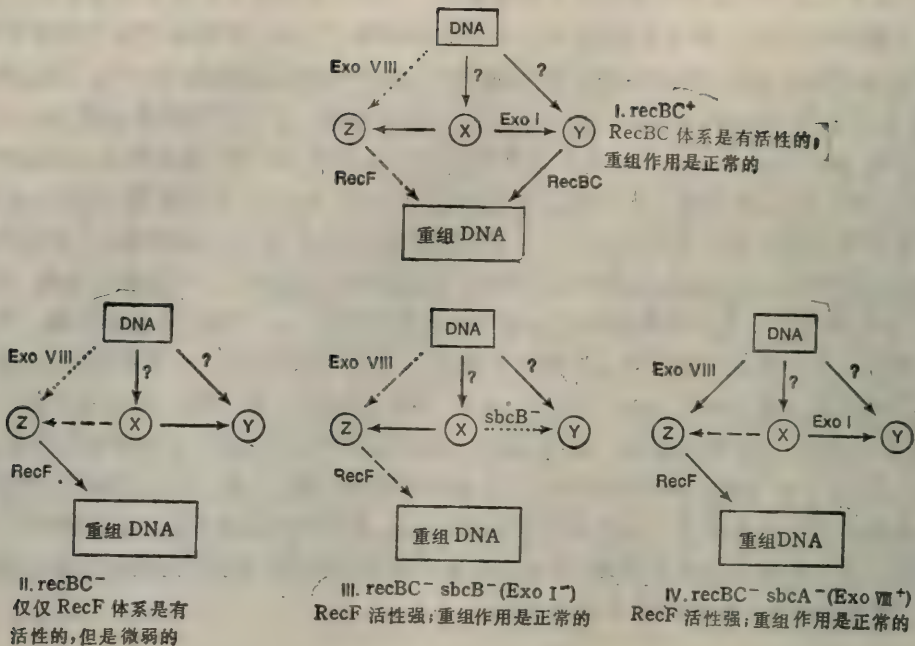


图 18-23 短虚线箭头表示并不发生的反应，长虚线箭头表示很弱的反应。问号表示对相应于这一步反应的酶尚不了解。

通过非专一性的途径形成 Y 或由  $sbcB$  产物将 X 转变成 Y。因此， $recBC^- sbcB^+$  突变体的重组活性只相当于野生型的重组活性的 1%，因为 RecBC 途径不起作用。RecF 途径需要中间物 Z，Z 可以由 X 通过非专一性途径转变而成。只要外切酶 I 具有活性，大部分 X 被转变成 Y，少部分 X 形成很少的 Z 去参与 RecF 途径，很低的重组频率 (1%) 反映了由 X 转变成的少量的 Z 所参与的 RecF 蛋白的活性。由于  $recBC^- sbcB^-$  缺乏外切酶活性，X 就不可能转变成 Y，全都变成 Z。所以  $recBC^- sbcB^-$  突变体表现出正常的 RecF 重组途径。中间物 Z 也可利用外切酶 VIII 来形成，这个酶是  $recE$  基因的产物， $recE$  基因的活性能被  $sbcA$  基因抑制。因此， $sbcA^- sbcB^+ recBC^-$  突变体也可以表现出正常的重组率，这是因为 Z 可以通过外切酶 VIII 的作用直接形成，而不必来自于 X；这时的 RecF 途径是有全部活性的途径。当然， $sbcA^- recE^- sbcB^+ recBC^-$  突变体是重组缺陷型的突变体 (又是只有 1% 重组率)，这是因为在外切酶 I 存在下，X 很少转变成 Z。如果  $recF$  的活性也被除去了，就不再有重组途径，因此， $recBC^- recF^-$  与  $recA^-$  一样，完全是缺陷型的，重组率极低。这个体系及其相互作用见图 18-23 所示，都很复杂。然而这里还有更多的几个  $rec$  基因 (从  $recJ$  到  $recM$ ) 未表示出来。 $recJ$  到  $recM$  基因的产物与前面所提到的基因的产物之间的关系还不清楚。

## 参 考 文 献

- Alberts, B., and C. F. Fox (eds.) 1980. *Mechanistic Studies of DNA Replication and Recombination* Vol. 19. ICN-UCLA Symposium on Molecular and Cellular Biology. Academic Press.
- Cunningham, R. P., T. Shibata, C. Das Gupta, and C. M. Radding 1979. "Single strands induce RecA protein to



- unwind duplex DNA for homologous pairing." *Nature*, 281, 191—195.
- Eisenstark, A., 1977. "Genetic recombination in bacteria." *Ann. Rev. Genetics*, 11, 369—396.
- Fox, M. S. 1978. "Some features of genetic recombination in procaryotes." *Ann. Rev. Genetics*, 12, 47—68.
- Holliday, R. 1964. "A mechanism for gene conversion in fungi." *Genet. Res.*, 5, 282—304.
- Hotchkiss, R. D. 1974. "Models of general recombination." *Ann. Rev. Microbiol.*, 28, 445—468.
- Lacks, S. 1962. "Molecular fate of DNA in genetic transformation in *Pneumococcus*." *J. Mol. Biol.*, 5, 119—131.
- McEntee, K., G. M. Weinstock, and I. R. Lehman. 1979. "Initiation of genetic recombination catalyzed *in vitro* by the RecA protein of *E. coli*." *Proc. Nat. Acad. Sci.*, 76, 2615—2619.
- Meselson, M., and C. M. Radding. 1975. "A general model for genetic recombination." *Proc. Nat. Acad. Sci.*, 72, 358—361.
- Potter, H., and D. Dressler, 1976. "On the mechanism of genetic recombination: electron-microscopic observation of recombination intermediates." *Proc. Nat. Acad. Sci.*, 73, 3000—3004.
- Radding, C. M. 1978. "Genetic recombination: strand transfer and mismatch repair." *Ann. Rev. Biochem.*, 47, 847—880.
- Sigal, N., and B. Alberts, 1972. "Genetic recombination: the nature of a crossed-strand exchange between two homologous DNA molecules." *J. Mol. Biol.*, 71, 789—793.
- Smith, H. O., D. B. Danner, and R. A. Deich. 1981. "Genetic transformation." *Ann. Rev. Biochem.*, 50, 41—68.
- Stahl, F. W. 1979. *Genetic Recombination*. W. H. Freeman and Co.
- Stahl, F. W. 1979. "Specific sites in generalized recombination." *Ann. Rev. Genetics*, 13, 7—24.
- Valenzuela, M., and R. B. Inman. 1975. "Visualization of a novel junction in bacteriophage lambda DNA." *Proc. Nat. Acad. Sci.*, 72, 3024—3028.
- Weissberg, R. A., and S. Adhya, 1977. "Illegitimate recombination in bacteria and bacteriophage." *Ann. Rev. Genetics*, 11, 451—473.
- West, S. C., E. Cassuto, and P. Howard-Flanders. 1981. "Homologous pairing can occur before DNA strand separation in generalized genetic recombination." *Nature*, 290, 29—33.
- Wilson, J. H. 1979. "Nick-free formation of reciprocal heteroduplexes: a simple solution to the topological problem." *Proc. Nat. Acad. Sci.*, 76, 3641—3645.

## 第十九章 遗传物质重排与交换机理 (III)

### 可转位因子

在前一章我们讨论了能在同源 DNA 序列发生交换的重组体系的性质。在细菌中,这种形式的交换依赖于 *recA* 基因产物或其等效物。在第十六章我们讨论了依靠处于同源序列两翼的伸展的非同源性识别序列而进行的短的同源区域的交换作用;这些交换作用要利用位点专一的重组酶,例如  $\lambda$  噬菌体的整合酶。在这一章中,我们将讨论另一种重组类型,叫作转位 (transposition)。在这个过程中经常可检测到出现的一段称为可转位因子 (transposable element) 的线形 DNA 片段。这种可转位因子本来就存在于细胞中,而且可在同一细胞的 DNA 分子上的一个第二位点上。可转位因子这个新的插入位点是一段与可转位因子没有同源性的碱基序列,而且这个插入过程不依赖于细菌 *recA* 基因。在早期的研究中,人们认为出现的可转位因子可以从一个位点移动或转位到另一个位点上。因此将这个过程称作转位。但是后来发现这个命名是错误的,因为在转位的过程中,可转位因子的一个拷贝仍留在原来的位点上,同时在新位点上出现第二个拷贝。因此转位与同源重组不同,转位过程必须进行 DNA 复制,更进一步,这种重组的唯一可能的后果就是可转位因子的移动。

与同源重组不同,转位作用通常是低频发生的。但它仍然是很有意义的,这不仅是因为它能说明在细菌中发现的许多基因缺失与基因倒转现象,而且对研究人员来说它是创造突变体的很有价值的方法。

### 几个术语

可转位因子通常也被称作转座子 (transposons)。有一些转座子带有易于识别的细菌基因,特别是那些抗生素的抗性基因。由于历史的原因,许多转座子用后面跟着一个数字的简略符号 Tn 来表示 (如 Tn5)。这个数目是为了与其他转座子区别。按国际的标准,一个新发现的细菌的转座子,即使它不带有标记基因,也应以此方式命名。当我们需要涉及到转座子携带的基因时,可以用标准的基因型表示方法来表示。如 Tn1(amp<sup>+</sup>),其中 amp<sup>+</sup> 表示转座子带有对氨苄青霉素的抗性基因。最初发现的转座子不含有任何已知的宿主基因;由于这个历史原因,又把它们称作插入序列 (insertion sequences) 或称作 IS 因子 (IS element),用 IS1 或 IS2 等来表示。有时,转座子也用不甚标准的方法来表示,如 F 质粒中的转座子就表示成  $\gamma\delta$ 。对于真核系统中的转座子还没有一个统一的系统命名,果蝇中的一个转座子称为 copia,另一个叫做 497,等等。

某个转座子常常被定位到一个特定基因中,这样就形成了这个基因的突变型。用标准命名法对这些突变可给以编号。下面的命名法用来表示由 Tn4 产生的 galT 基因的 135 号突变:



注意使用了两个冒号。

## IS 序 列

1967年,人们第一次观察到的可转位因子就是 IS 序列。发现 IS 序列的过程十分有趣。现在已经知道 IS 序列只是转座子的一种,然而可以很容易地通过对 IS 序列这个特殊类别进行仔细观察,而看到转座子的某些极为重要的性质和各种各样的检测转座子存在的方法。

### IS 序列的发现

人们在分离大肠杆菌的 *gal* 操纵子和 *lac* 操纵子等一类高度极性的突变株时,首次发现了 IS 序列。*gal* 和 *lac* 突变都位于操纵子的第一个结构基因中,突变点下游不能产生所编码的蛋白产物了。但是如果这个细菌含有 *rho* (*ts*) 突变(生长对温度敏感的突变),那么这种极性突变就可以被抵销。很显然,极性突变依赖于 Rho 蛋白的终止序列的存在。这些极性突变的另一个特点就是它们不能被碱基类似物或移码突变逆转成野生型,因此,极性突变不可能是由于碱基突变造成的。对含有极性突变的质粒进行物理图谱分析,表明极性突变的操纵子中含有一段插入的 DNA 片段。在几个噬菌体系统中(如 T4 和  $\lambda$ ),明显的插入作用是由于复制错误造成的。复制错误产生了碱基序列的串联复制。但是几个方面的证据证明插入的片段并不是简单地由邻近序列的复制所形成的。因此,这段插入序列的来源还不完全清楚。有假说认为,它们是能从一个染色体上转移到另一个染色体上的可移动的序列。下面将简单地说明它们的可移动性。

### IS DNA 的大小及性质

用电子显微镜观察携带着上面介绍的极性突变的特化的转导噬菌体颗粒。将从携带着突变的转导颗粒中提取出来的 DNA 与从没有突变的转导颗粒中提取出来的 DNA 进行分子杂交,杂交后出现一个泡(见图 19-1),这个泡表明该突变是由一个长度等于一个泡的插入片段引起的。比较不同的含有极性突变的转导颗粒,指出一个颗粒的 DNA 泡可以和另一个颗粒的 DNA 泡杂交,这表明在每一个突变的位点中存在着相同的序

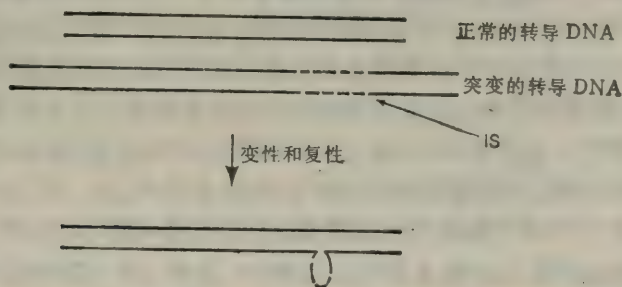


图 19-1 示出 IS 序列是一段插入 DNA 片段的杂交实验。在复性后,有一半的分子已再度形成正常的和突变的转导分子(图中仅示出杂交双链分子)。

列。

从插入了 IS 因子的大肠杆菌染色体的各个区域提取得到许多基因,把它们转到噬菌体的转导颗粒上;又将插入序列引入到某些质粒上。通过对含有上述基因的大量的转导颗粒以及含有插入序列的质粒进行研究,发现有好几种不同的 IS 因子。这些因子中的某一些已经从转导颗粒的 DNA 中分离出来,并测定了它们的碱基序列。那些能产生极性突变的片段都含有依赖于 Rho 的转录终止信号(不包括 IS1 的),并在所有可能的阅读框架

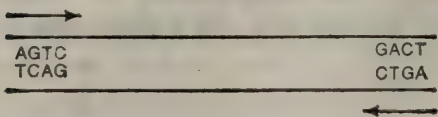


图 19-2 DNA 分子中末端倒转重复的一个例子。箭头表示倒转重复序列。注意,序列 AGTC 和 CTGA 不在同一条链上。在一种所谓直接重复的情况下,上面一条链中的 DNA 序列应当是 AGTC.....AGTC。

中都含有链终止突变信号,这就说明了极性效应。一个对 IS 因子的碱基序列的重要发现是观察到此因子的末端有倒转重复序列(见图 19-2)。这些倒转重复序列长度为 16—41 碱基对,具体究竟多长取决于转座子。在许多情况下(例如大肠杆菌的 *gal* 操纵子中),插入序列既可以按左→右,也可以按右→左的方向指到插入位点上。人们可能认为这样的

情况发生在所有的插入位点上,因为倒转重复序列是对称的;但是,实际上还观察到存在着许多只能按一个方向插入的位点。少数 IS 因子的某些更为重要的性质列于表 19-1 中。

表 19-1 某些大肠杆菌插入因子的性质

因 子	位置及拷贝数	大小,碱基对	碱基序列数据
IS1	染色体中 5—8 个	768	完全的
IS2	染色体中 5 个 F 中 1 个	1 327	完全的
IS3	染色体中 5 个 F 中 1 个	约 1 400	几乎完全的
IS4	染色体中 1~2 个	约 1 400	仅仅有末端的
IS5	不清楚	1 250	仅仅有末端的
$\gamma\delta$	染色体 1 个或多个 F 中 1 个	5 700	仅仅有末端的

IS 序列在大肠杆菌 DNA 和一些噬菌体 DNA 中的位置

在前一节中我们暗示,引起极性插入的 IS 序列是来自细菌,事实上,像表 19-1 指出的那样,确实如此。大肠杆菌 DNA 上存在着许多个 IS 序列的拷贝。通过杂交试验测出了每个 IS 序列的拷贝数。它们在大肠杆菌染色体中的位置已经通过电子显微镜检查得到确定。人们从质粒或带有大肠杆菌染色体特殊区域的特化的转导噬菌体中提取到 DNA,再用电镜观察那些自我退火的大肠杆菌 DNA 双链或大肠杆菌 DNA 与质粒 DNA 及与转导噬菌体 DNA 杂交的双链,确定了 IS 序列的所在位置。例如,有两个 IS 序列位于一个 *arg* 基因的两侧。电镜观察表明质粒也含有 IS 序列,如 IS2、IS3 和  $\gamma\delta$  就存在于 F 因子中,而且位于 F 因子整合到大肠杆菌染色体形成 Hfr 的位点上(见图 17-11 和 17-12,第十七章)。而且,在一些 R 质粒中也有 IS 序列,它们位于抗药性基因与那些负责复制及结合转移 (conjugal transfer) 基因(图 17-15)之间的区域的边缘。在质粒中的这些位点的重要性将在以后讨论。



IS 序列也存在于大肠杆菌的几种噬菌体中,例如 IS1 是噬菌体 P1 的一个正常组成部分;在  $\lambda$  噬菌体的一些特殊的突变株中也有 IS,而且这些各种各样的 IS 序列已被定位。在不同的生物中还发现了一些引人注意的与之相匹配的序列,例如,有一个在 IS1 序列两端的十二核苷酸序列,它与  $\lambda$  前噬菌体整合位点的十二核苷酸序列相同;并且发现 IS1 的靠近末端处的序列,类似于  $\lambda$  的左向启动子 (pL) 和 tRNA<sup>Tyr</sup> 的启动子的序列。以上这些发现的重要性现在还不很清楚。

# 转 座 子

IS 序列是转座子的一种特殊种类,在这部分中,我们将讨论转座子更多的一般特性以及它们被发现的过程。

## 转座子的发现

在前一节中,我们介绍了由于 IS 因子在基因上产生的极性效应而使我们能够用生物学方法正常地检测到它们。但是对于典型的细菌转座子并不经常使用那些方法。因为转座子经常含有可识别的细菌基因,例如,对抗菌素的抗性基因,通常通过这些基因我们就可以更简便地检测到转座子的存在。我们将讨论两个例子,如图 19-3 和 19-4 所示。

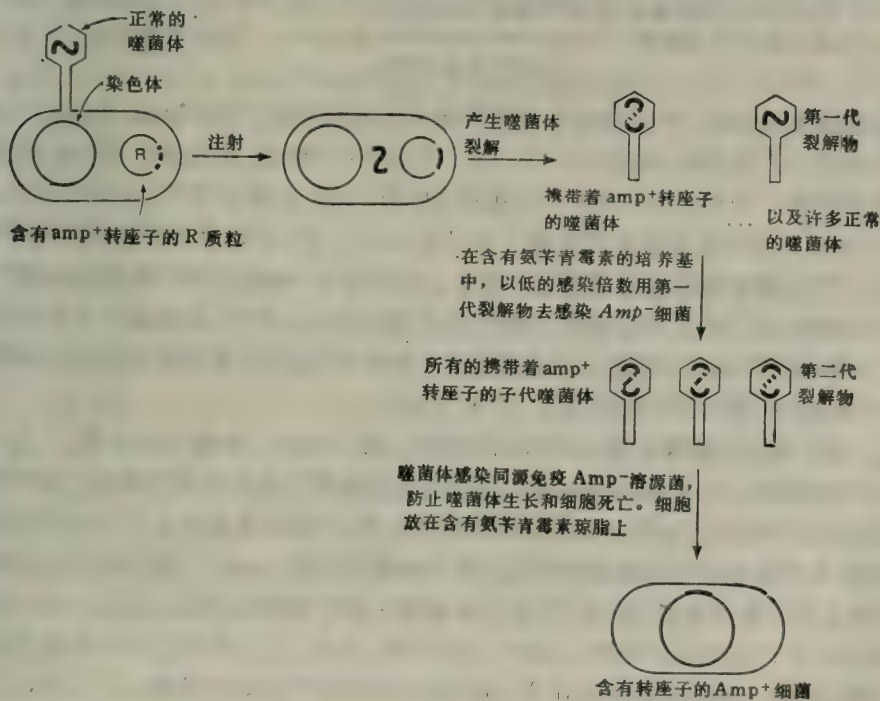


图 19-3 证明转座作用的实验。用虚线表示转座子,细节见正文。

图19-3描述了检测转座子的过程。用温和噬菌体去感染一种同时又带有  $amp^+$  的 R 质粒的  $recA^-$  细菌,而且这个  $amp^+$  基因位于转座子上。这样,噬菌体就可获得  $amp^+$  基因。在这个实验中运用  $recA^-$  细菌是为了消除一切同源重组的可能。在这个实验中,

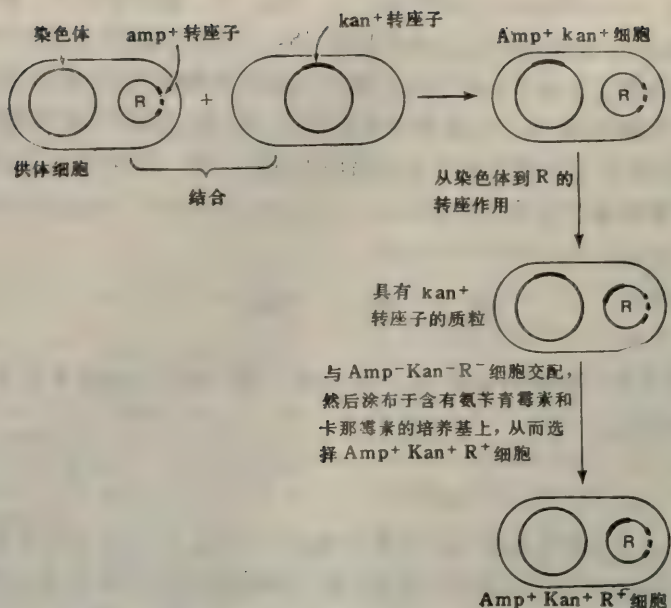


图 19-4 证明转座作用的实验。含有  $\text{amp}^+$  R 质粒的供体菌将 R 质粒转移到含有  $\text{kan}^+$  转座子的受体菌中去。在下一代中  $\text{kan}^+$  转座子又转移到 R 质粒上。R 质粒一次可以转移  $\text{amp}^+$  与  $\text{kan}^+$  两种抗药性基因的现象，证实了 R 质粒上存在着第二个 R 质粒。用  $\text{amp}^+$  基因表示整个转座子就像大部分的抗药性决定子 (drug-resistance determinates)。但这个实验对于证明转座作用并不是必需的。

当细菌被噬菌体感染后，一小部分子代噬菌体 ( $1/10^7$ ) 就含有了带  $\text{amp}^+$  基因的转座子。将得到的噬菌体以噬菌体:细菌为 1:10 的比率去感染  $\text{Amp}^-$  的细菌，同时在培养基中加入氨苄青霉素。培养皿中未受感染的细菌 (仍是  $\text{Amp}^-$ ) 都被杀死。更重要的是，不带有  $\text{amp}^+$  基因的噬菌体所感染的细菌 (也仍是  $\text{Amp}^-$  的) 不能生长繁殖，也不会产生子代噬菌体，因为被氨苄青霉素杀死了。但是，如果带有  $\text{amp}^+$  基因的噬菌体去感染细菌时，细菌就会存活下来，产生带有  $\text{amp}^+$  的子代噬菌体。这样，在有氨苄青霉素存在的条件下，只有带有  $\text{amp}^+$  转座子的  $\lambda$  噬菌体才能够生长；因此受感染细菌的裂解物是由同源的  $\lambda\text{amp}^+$  噬菌体群体组成的。

为了证明  $\text{amp}^+$  基因是转座子的一个部分，用  $\lambda\text{amp}^+$  噬菌体去感染一个  $\text{amp}^- \text{recA}^-$  的  $\lambda$  溶源菌。由于溶源化细菌对造成溶源化的噬菌体有同源免疫性，即对同源的噬菌体的侵入具有抵抗力，所以噬菌体不能发育扩增，于是细菌能生存下去。然后将被感染的细菌涂布在含有氨苄青霉素的琼脂培养基上进行培养。 $\text{amp}^+$  噬菌体使细胞仍然能够生长，但是噬菌体本身却不能复制 (由于  $\lambda$  噬菌体的阻遏物的存在)，所以在每一次细胞分裂周期中，子代细胞不能得到  $\lambda\text{amp}^+$  的 DNA 分子，因此细菌也不能形成菌落。然而，在少数细菌中，确实发生转座作用，细菌染色体获得了  $\text{amp}^+$  转座子。因此，这种子细胞也确实都获得了  $\text{amp}^+$  基因。于是可以见到细菌菌落的形成。这样，通过能够形成菌落也就说明发生了转座作用。注意，这样产生的  $\text{amp}^+$  细菌能被正常噬菌体再度重新感染，产生第二个 (即新的) 噬菌体的群体，其中少数带有转座子。

人们可以利用这种方法来产生带有特殊基因的突变株。例如，一个转座子可以插到



ac 基因中。这样,如果上图的最后阶段用的  $\text{amp}^-$  溶源菌是  $\text{Lac}^+$ ,而我们想要得到  $\text{Lac}^-$ ,那么我们可以在几百个  $\text{Amp}^+$  的菌落中找到一个  $\text{Amp}^+ \text{Lac}^-$  的菌落。这个工作可以通过下列步骤很容易地做到,那就是把受感染的细菌培养在含有乳糖和氨苄青霉素的、有带色指示剂的琼脂培养基上;通过不同的颜色即可将  $\text{Lac}^+$  菌落和  $\text{Lac}^-$  菌落区分开来。这种方法是用来制备缺陷型突变株的一个很有用的技术。因为用这种方法使细菌中产生某一个特定基因发生突变的抗药性菌落的比例要高于用大部分突变剂诱导产生的突变型细胞的比例。因为编码序列中插入了转座子,所以,如此形成的突变当然是绝对的缺陷型突变,而不是条件性突变。

在图 19-4 中所示的第二个实验中,含  $\text{amp}^+$  转座子的 R 质粒通过结合,转移到含有  $\text{kan}^+$  转座子的  $\text{recA}^-$  受体菌中。在两种抗生素都存在的情况下继续生长,偶尔会发生转座作用, R 质粒带有两种抗药性标记及两个转座子(每一个转座子含有一种抗药性基因)。通过结合,这种双转座子的 R 质粒能转入到  $\text{Kan}^- \text{Amp}^-$  细菌中,并生长在含有双抗生素的培养基上,这样就证实了带有双转座子的 R 质粒的存在。如果需要的话,在培养的过程还能再次观察到转座作用,因为在培养很多世代以后,其中某种转座子(例如  $\text{amp}^+$  转座子)还可以转座到染色体上。如果这种情况确实发生了,那么用遗传学方法从细胞中移去质粒(如吡啶疗法 acridine curing, 见十七章)后,细胞仍保持  $\text{Amp}^+$ 。这种在染色体与 R 质粒以及在两个抗药性质粒之间的转座作用是形成一些具有多种抗药性的质粒的主要途径。在医院里,对受感染的病人施给药物,有时会遇到多重抗药性反应,这就是由上述质粒引起的。

在图 19-3 与图 19-4 所示的实验中,含有两种抗药性的噬菌体或质粒可以与原来的噬菌体或质粒杂交,并且可以用电镜检验。从图 19-1 中我们可以看到一些泡,这些泡就表明存在着插入的 DNA 序列。

### 转座子插入位点的定位

当发生转座作用时,通常一个特殊的转座子 Tn 可以在很多地方发生插入。例如,当一个带有可转座的抗生素抗性标记的噬菌体去侵染野生型  $\text{recA}^-$  大肠杆菌时,就会像图 19-3 下部所示的那样,人们通过检验到产生了抗药性细菌而检测到有转座事件发生。例如,在被感染的  $\text{lac}^+ \text{leu}^+ \text{amp}^-$  细菌培养物中可以产生出  $\text{lac}^- \text{amp}^+$  或  $\text{leu}^- \text{amp}^+$  的突变体。如果检验上千个  $\text{Amp}^+$  细菌菌落,通过测试细菌所需要的营养成分以及利用某种糖作为碳源的能力,结果发现在大部分带有大转座子的菌落内,几乎任何的基因都会发生突变;只是在不同的菌落中,突变的基因不一样,所有突变体的突变基因的总和囊括了大肠杆菌的所有基因。这说明对于大转座子的转座作用来说,插入位点很多,散布在整个染色体上。但是,对一些小的转座子(例如 IS 因子)来说,插入位点的数目则看来似乎是有限的。

另有一个类似的实验说明即使在很小的 DNA 片段(分子量约为  $10 \times 10^6$ )中,转座子的插入位点也是很多的。例如,在利用 P22 噬菌体做的实验中(如图 19-3 上半部分所示),能分离到带有抗药性标记的、许多不同的 P22 突变体。为了确定插入位点,先从这些噬菌体中分离出 DNA,将这些 DNA 分别地与野生型 P22 的 DNA 进行杂交,然后再在电镜下观察泡(loop)的位置(图 19-1),泡说明了插入位点的位置。而且

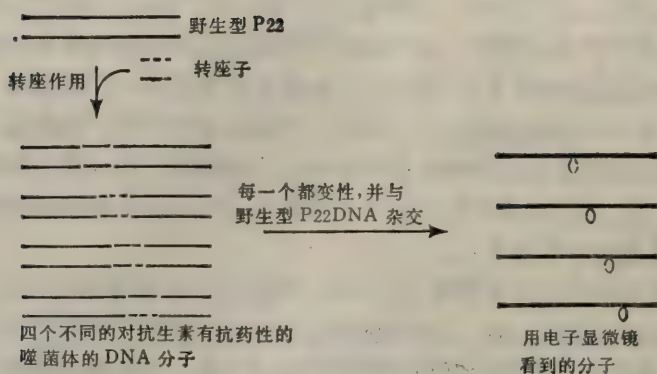


图 19-5 说明特定转座子可插入到 P22 DNA 的不同位点上的一个实验。

如图 19-5 所示,在许多位点上都可发现存在着泡状结构。

虽然已经确定在 P22 与  $\lambda$  中必定有许多插入位点,但是很显然,这些位点也并不是随机分布的。有一些是不易被插入的,而另一个则可重复地使用(这些位点被称为热点 hot spots)。检验已知序列的很小区域中的插入位点就可以看到上述现象。这种研究表明从一个转座子到另一个转座子的插入作用专一性的范围很宽。一个极端的例子就是 IS4,人们观察到它在 gal T 基因中的 20 个独立的插入事件,所有这些插入事件多半是插到具有同样的核苷酸序列的位点上。另一个极端的例子是 Mu 噬菌体, Mu 几乎能在任何地方插入,这将在以后再讨论。大多数转座子显示的插入作用专一性则是介于上述两个极端例子之间。例如,转座子 Tn9 在 lac Z 基因的最后 160 碱基对中的插入位点的序列已经测出;总共有 28 种插入作用,插入位点则有 16 个,其中 5 个是多次发生插入的地方。转座子 Tn 10 更特别,人们已注意到它在鼠伤寒沙门氏杆菌的 his 操纵子中有 22 个不同的插入位点,但是已知的转座作用约有 40% 都发生在同一个核苷酸序列的位置上。对于引起插入的热点则知道得还很少。

### 转座子的类型

转座子可以分成三种类型:(1) IS 序列以及含 IS 的转座子;(2) Tn3 类;(3) Mu 噬菌体类。这些类型有以下性质:

1. IS 序列和复合式转座子。已讨论的 IS 序列为长度 750—1 500 碱基对的 DNA 小片段(见表 19-1)。遗传学分析以及对碱基序列中起始密码与终止密码的鉴定指出 IS 序列至少编码一种蛋白质,而这种蛋白质对于转座作用是必不可少的。一些 IS 序列还能编码第二种蛋白质,这种蛋白质很可能起着调节作用。

一种携带着一些抗药性基因的转座子的变种,是由那种两翼具有两个相同的或近乎相同的 IS 序列拷贝的可转移因子所组成,其中的抗药性基因本身也是其他转座子的组成成分。这些转座子被称为复合式(composite)类型 I 转座子。在这些复合式转座子单位中的 IS 序列可以是倒转重复的构型,也可以是正向重复的构型(图 19-6)。因为 IS 的两个末端是它们本身的倒转重复,所以复合式转座子的两翼的相对方向(反向或正向)并不改变它们的末端序列;也就是说,在倒转重复的排列中它们仍然保持着相同的序列。



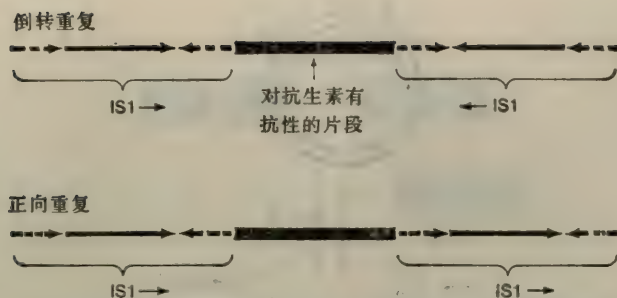


图 19-6 两侧带有倒转重复或正向重复的  $IS1$  序列的两个复合式类型 I 转座子。

在大多数情况下,转座子的转座能力是由  $IS$  序列决定并调节的,表 19-2 列出了几种经过较详细研究的复合式转座子和它们的一些性质。

表 19-2 大肠杆菌的几种复合式类型 I 转座子的性质

因 子	携带的基因*	大小 (bp)	末端的 $IS$ 因子及其 大小 (bp)	末端的 $IS$ 因子的相 关方向
Tn5	kan	5 400	IS50 (1 500)	反向
Tn9	cam	2 638	IS1 (768)	正向
Tn10	tet	9 300	IS10 (1 400)	反向
Tn204	cam, fus	2 457	IS1 (768)	正向
Tn903	kan	3 100	IS903 (1 050)	反向
Tn1681	ent	2 088	IS1 (768)	反向

\* cam, 氯霉素; ent, 肠毒素; fus, 梭链孢酸; kan, 卡那霉素; tet, 四环素。

绝大多数复合式类型 I 转座子是在它由一个质粒转座到另一个质粒时或转座到噬菌体时首次作为遗传因子而被观察到的。当这些质粒 DNA 被变性,然后再复性时,在电泳下观察到带有柄与泡结构的单链环。如图 19-7 所示,这样的形状表明倒转重复序列的存在。用这些质粒与其他已经带有已知  $IS$  序列(例如  $IS1$ )的分子进行杂交分析指出,其中有一些质粒确实含有  $IS$  序列。用这种方法说明了 Tn9(表 19-2)有两个含倒转重复的  $IS1$  序列。其他的不具有能与已知  $IS$  因子杂交的序列的质粒,由于其柄与泡结构中的柄差不多都一样大,所以人们设想这些转座子可能也有由尚未鉴定的  $IS$  因子所组成的两翼。在退火之后也可能见到双链环。这些双链环中含有由 DNA 中的同一位点而产生的两个柄与泡结构,这种 DNA 也是由倒转重复序列组成的。另外的杂交实验表明,还有一些含已知  $IS$  序列的质粒并不形成柄与泡的结构。事实证明其中的  $IS$  序列是正向重复的。

2. Tn3 转座子家族。Tn3 家族转座子由一些非常大的(约 5 000 碱基对)转座子组成。每个转座子带有三个基因,其中一个编码  $\beta$ -内酰胺酶(它提供出 amp 抗性),另两个是转座作用所必须的(将在以后讨论)。所有的类-Tn3 的转座子含有短的(38 碱基对)倒转重复,而且两翼不具有  $IS$  序列。最近有人认为以前曾被认为是 Tn3 家族的转座子(例如 Tn501, Tn551, Tn1721 和  $r\delta$ )可能属于第四类型。

3. 可转座的噬菌体 (Mu 类)。两种有关的噬菌体 Mu 与 D108 把转座作用作为它们产生子代噬菌体的正常方式的一部分。在感染的早期,这些噬菌体将他们的 DNA

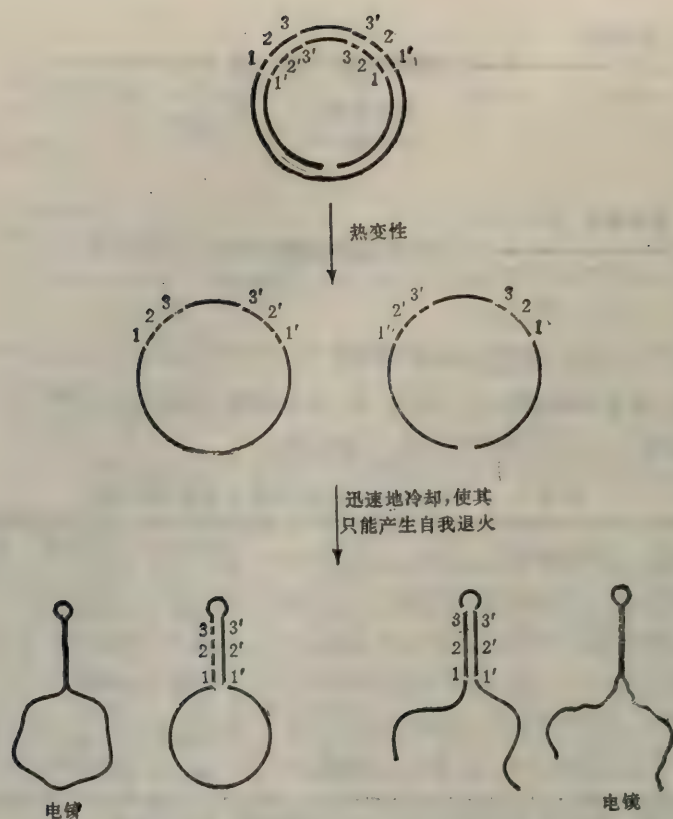


图 19-7 含有倒转重复的质粒经变性及自我复性形成柄与泡 (stem-and-loop) 结构。标记着电镜两字的构型是在电镜下看到的情况。细线、粗线分别表示单、双链。

整合到染色体上 (染色体上有许多个整合位点, 这个整合先发生在其中一点上), 产生典型的转座子序列, 即噬菌体 DNA 夹在两个复制了的靶序列 (作为侧翼) 之间。按转座作用普遍存在着的原理来看, 噬菌体 DNA 的多个拷贝转座到染色体 DNA 上的许多位点上, 最终由这些染色体上的转座单位进行噬菌体包装。正像在第十六章所解释的那样, 包装要将噬菌体的头壳装满, 包装进去的 DNA 长度大于噬菌体的编码序列; 因此, 噬菌体颗粒包装进去的 DNA 的末端总含有细菌 DNA。

## 转 座 作 用

转座过程的最终结果是受体 DNA 分子的两个碱基序列中间插入了转座子。对许多转座子及他们的插入位点进行序列分析, 说明它们之间没有序列的同源性。这与它们不需要大肠杆菌 *recA* 系统是一致的。然而受体 DNA 的插入位点区域的序列很令人吃惊, 我们将在下一节加以讨论。

### 插入位点的靶序列的复制

转座时发生的插入作用有一个普遍的特征, 那就是在受体 DNA 分子中有一段很短的碱基序列 (3—12 碱基对), 人们称它们为靶序列, 它会进行复制。插入的转座子就被夹



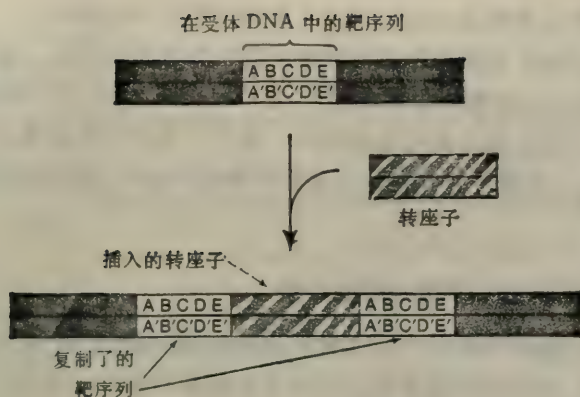


图 19-8 转座子的插入造成靶序列的复制。

在两个重复着的靶序列之间,这种排列见图 19-8。我们再强调一下,在转座子插入之前,受体 DNA 中只有一个拷贝的待复制的靶序列,转座子本身不存在这种靶序列。

不同转座子的靶序列的长度不同,但是对于某一个转座子来说,它的所有的靶序列的长度都是一样的。例如,IS1 的两翼总是有 9 碱基对的一段靶序列,而 Tn 3 的两翼总是 5 碱基对的序列。注意,对于某种特定的转座子来说,每一种插入位点的靶序列都不同,只是复制的序列的长度是恒定的。

下面对负责靶序列复制的机理作一些介绍,但其细节还不清楚,在图 19-9 只表示了大概的过程。

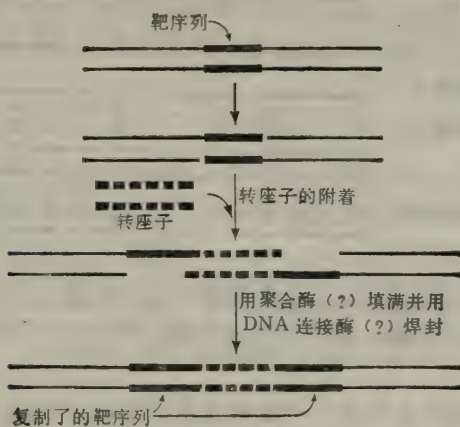


图 19-9 靶序列复制过程图解。

一个可能由转座子编码的酶能够识别与选择受体 DNA 中的靶序列(其选择方式还不清楚),并在两条链上的靶序列的末端各形成一个单链缺刻。然后,转座子 DNA 附着到由缺刻形成的自由末端上,因此转座子中的一条链与靶序列的一条链相连,另一条链与靶序列的另一条链在相对的那一头的末端相连。这种相连会造成两个缺口,每一个缺口跨在靶序列的一条链上,然后,缺口被填补起来(可能是 DNA 聚合酶 I 的作用),进一步缺刻被焊封起来(可能是 DNA 连接酶的作用);至此,形成了连续的双链的受体 DNA,而且,在转座子两翼有两个靶序列。注意,当转座子与受体 DNA 结合时,并未利

用末端的同源性。

在本章的后面部分将更详细地说明产生复制的靶序列的机理。

因为每个整合了的转座子都被夹在两个复制过的靶序列中间，甚至那些转座子呆在固定位点时也是这样（如 F 质粒中的转座子）。人们不禁要问，这种复制作用对于从供体位点起的转座作用来说复制是否也是必需的？事实上肯定不是必需的，因为已经构建出（用重组 DNA 技术）两翼没有直接连着重复序列的转座子，这些转座子仍然能够发生转座作用。

转座子的结构和转座子-靶序列复合体的性质

已用限制性内切酶对含转座子的 DNA 进行切割，从而分离到了含有转座子的 DNA 片段（见第二十章）。靶序列以及相邻的转座子的碱基顺序都已测出。通过这些分析发现了两个重要事实：

1. 对于一个特定的转座子的每一个插入作用来说，与两个靶序列为邻的转座序列都是一样的（图 19-10）。每一种转座子都有自己专一的序列。这说明每一种转座子都有严格确定的末端，并说明转座过程是通过转座子的末端与靶序列的相反一侧上的缺刻相互连接而完成的。

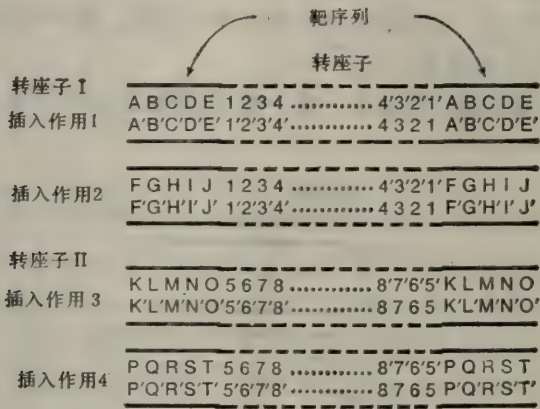


图 19-10 图示同一个转座子在两个不同插入位点插入时，转座子与靶序列连接处的序列。即 1234 是转座子 I 在插到所有插入位点时的左端序列，转座子 II 插入时的左端序列则总是 5678。每个数字及字母都代表一个碱基。带“'”的表示互补碱基。数字与字母只是一种符号；5678 表示与 1234 不同的序列，而不是跟在 1234 后面的序列。

2. 上述的转座子的两个末端由一种倒转重复的序列排列而成（可发生转座的 Mu 噬菌体除外，后面将简单地讨论 Mu 噬菌体），如图 19-10 所示。注意，按 5'→3' 方向进行阅读的时候，转座子 I 的左上部分的序列与右下部分的序列都是 1234。因转座子不同，倒转重复序列的碱基数可以变动于 8—38，在图 19—11 中给出了其中几个序列。两个末端的相互对称并不像图中的直立箭头所示的那样完好。对 Mu 噬菌体来说，虽然它有许多更为复杂的部分是对称的，但是它的末端却不是倒转重复的。

许多遗传学证据说明转座子的末端序列是转座作用所必需的。例如，已经分离到一些发生了一个碱基对改变的、有小的缺失和小的插入的倒转重复序列。一旦出现以上的这些变化，转座作用就受到阻碍。



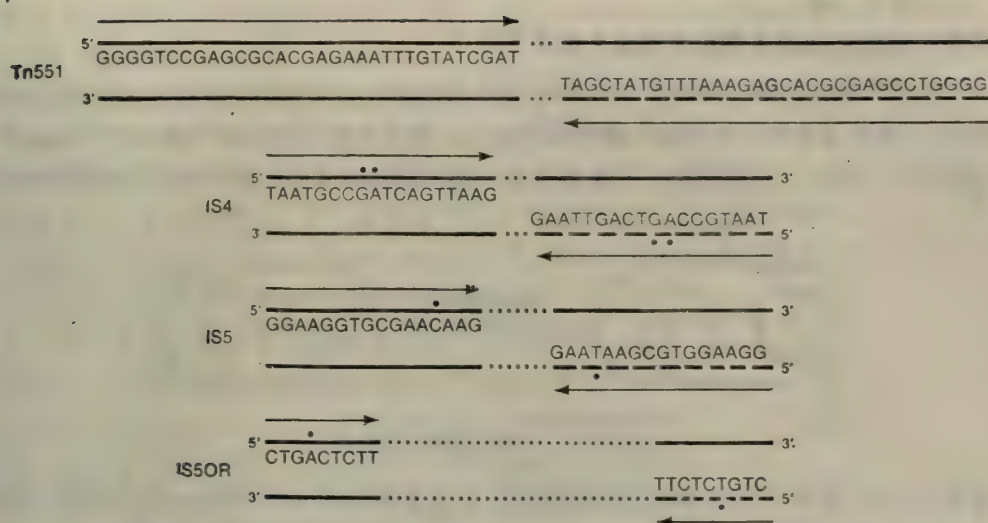


图 19-11 四个转座子的末端序列。图中表示了此种序列的倒转重复。除了由黑点表示出的个别情况以外,如果按箭头方向阅读,虚线表示的序列与实线表示的序列是一样的。

## 转座机理

至此,还未提到转座的分子机理,这主要是由于还不清楚这种机理。下面介绍由转座子介导的称作共整合的现象,将有可能使我们能更好地了解转座子。下面我们将要讨论在转座过程中假设的共整合的作用以及几个有关这种机理的假说。

### 共整合

首先考虑一个带有两个质粒的细胞,其中每一个都含有一个转座子。在每一个细胞世代中,两个质粒融合形成一个质粒的频率为  $10^{-7}$ ,即占细胞总数的千万分之一的细胞中才会发生上述现象,这个过程就称作共整合 (cointegration),有时也称作复制子融合 (replicon fusion)。这个杂交的质粒并不简单地就是两个质粒的融合,因为它有两个拷贝的转座子(而不是仅仅为一个)。这两个转座子拷贝的方向是相同的(即是正向重复的),而且它们严格地位于供体与受体质粒之间(图 19-12)。按目前所知,所有的转座子都可以介导共整合作用,虽然有时产生的共整合物并不稳定。

由于两个原因使共整合的形成很有意义。首先,人们认为共整合产物是转座过程必要的中间物;第二,在共整合前只有一个拷贝的转座子,而在共整合后却出现两个拷贝

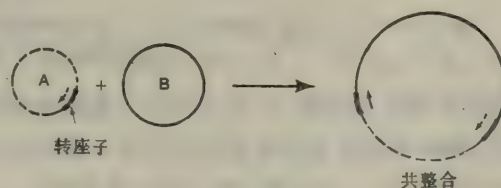


图 19-12 由转座子介导的 A、B 质粒融合而形成的共整合。在这个过程中产生两个拷贝的转座子。箭头是按任意方向画的,只是为了表明在共整合中两个转座子的方向是相同的。

的转座子的事实,清楚地说明了转座子 DNA 发生了复制过程。在下两节中将更进一

步讨论这两点。

在两个质粒之间发生转座时转座子的复制过程

在上一节中说明了一个含有一个转座子的质粒可以与第二个质粒融合，形成一个共整合物，即含有正向重复的两个拷贝的转座子。但是观察到的更多的是真正的转座作用，如图 19-13 所示。也就是说，一个含有带转座子的质粒 A 与不带转座子的质粒 B 经过许

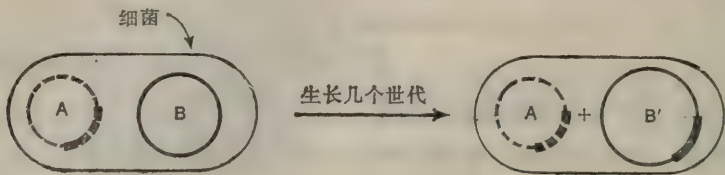


图 19-13 转座子从质粒 A 到质粒 B。

多世代以后，就会产生含有带转座子的质粒 B（也称作 B'）的细胞。其中重要的发现就是，如果在一个细胞中含有两个质粒，那么在两个质粒中就都有同样的转座子拷贝。这就说明在转座过程中原初的转座子序列发生了复制过程，就像前一节所叙述的那样，转座过程是一个复制的过程。

另一个很重要的发现就是，如果含有共整合物的细胞（当然已具有两个拷贝的转座子）生长很多世代以后，就会产生一些只各带有一个转座子拷贝的质粒的细胞，其中没有共整合物了，这再次暗示共整合作用可能是转座作用的第一个步骤。在下一节将再讨论这个观点。

在 Tn3 的转座作用中共整合物作为中间物存在

用 Tn3 转座子做的一些实验明显地表明，共整合物的形成是转座过程中的一个中间步骤。

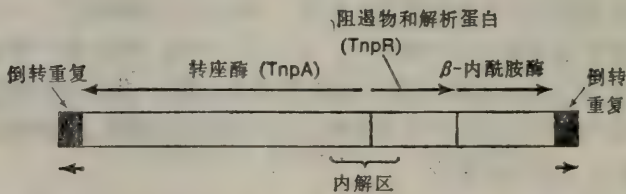


图 19-14 Tn3 转座子的物理图谱。共有 4 957 碱基对；每个倒转重复序列都含有 38 碱基对。在三个箭头的上面标明了三个基因。

Tn3 的结构在图 19-14 中给出。像所有的转座子一样，Tn3 有一对倒转重复序列，在这两个侧翼中间夹着三个基因。已经分离到这三个基因的产物，从它们的大小以及 Tn3 的序列可知，这三个基因利用了倒转重复序列之间的全部碱基，最左边的基因编码一个很大的蛋白质（1 015 个氨基酸）。这个蛋白质记作 TnpA，它是一个转座酶并负责形成共整合物。最右边的基因编码  $\beta$ -内酰胺酶，这个酶使氨苄青霉素失去活性。中间的基因编码一个小的蛋白质（185 个氨基酸），称作 TnpR。这个蛋白质具有两个功能，即负调控作用或者说是阻抑 TnpA 的合成，以及启动转座的第二步：专一性的位点交换，这种交换解开转



座过程的第二步中的共整合作用。在编码 TnpA 蛋白基因与编码 TnpR 蛋白基因的交界处的前后,是一段被称作“内部解体位点”(即内解区, internal resolution site) 的 DNA 序列, TnpR 蛋白作用于其中的一段序列。

遗传学实验得出了如下的结果:所有的 tnpA<sup>-</sup> 突变都不能转座或形成共整合物。而且, tnpR<sup>-</sup> 突变以及内解区中有缺失的形式只能形成共整合物但不能转座。因此,在一个含有两个质粒的细胞中,其中一个质粒带有 tnpR<sup>-</sup> 拷贝的 Tn3, 另一个不带 Tn3, 会发生图 19-12 所示的过程,而不会发生 19-13 所示的过程。这些发现说明 Tn3 介导的转座过程分为两步:首先, TnpA 诱导共整合,然后, TnpR 促进发生在内解位点的专一性位点交换。图 19-15 给出了这个过程的图解。

有一些转座子不带有类 TnpR (TnpR-like) 的专一性位点重组体系,但是仍然能发生转座。推测这些体系是利用了依赖于同源交换的体系,因为,一旦共整合发生,就有两个相同的序列,即转座子及其拷贝。所以任何依赖于同源序列的过程都能实现图 19-15 所示的那种过程。实验事实支持这种看法,那就是如果在宿主细胞中存在 recA 基因,缺失 TnR 或缺失内解区的 Tn3 也能实现图 19-13 所示的过程,即把共整合物转变成单独的质粒;事实上,在已经形成共整合物的 RecA<sup>-</sup> 细胞中,通过转导颗粒引入等位基因 recA<sup>+</sup>,就能发生转座作用。

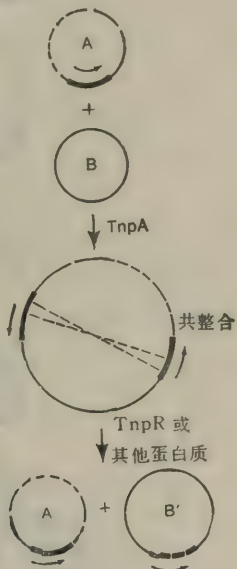


图 19-15 利用共整合物作为中间物的转座模型。

转座作用的两种模型

转座的机理还不清楚,但现在的说法已有很多。这些说法必须能解释下述事实:(1) 转座发生后,在原来转座子的位置上还保留着转座子,即转座过程是复制过程;(2) 共整合物的存在;(3) 产生短的重复靶序列 DNA,这种靶序列处在新整合进来的片段两侧;(4) 一些转座子需要有内解区,而另一些则依赖于同源性的交换。

已经提出过许多关于转座子的模型,其中有一些是大同小异的,另一些则差别很大,所有的模型都有两个共同的特点:

- 1. 两个单链被连到一起(是连接的),而不必靠碱基配对指出连接的位置。
- 2. 连接作用产生一个复制叉。

读者应当注意在下面两个模型中的这些特点。

图 19-16 给出了一个为人们普遍接受的模型。在第一个模型中,共整合物被认为是一个中间物。在第二个模型中,共整合物不被认为是一个中间物,只是转座子简单地插入,这第二个模型很可能解释不了 Tn3 的转座作用,只适用于其他那些没有内解区的转座子。

我们首先考虑带有一个转座子的供体质粒和一个受体质粒(供体和受体都不一定必须是质粒,甚至可以不是环形的 DNA。图中所示那种方式是研究得最多的排列方式)。在受体中,转座子发生整合的位点是一段短的碱基序列——靶序列。还无人提出关于转座子与靶序列如何连接的设想。有人建议转座步骤如下(文中的编号与图 19-16 相同)。

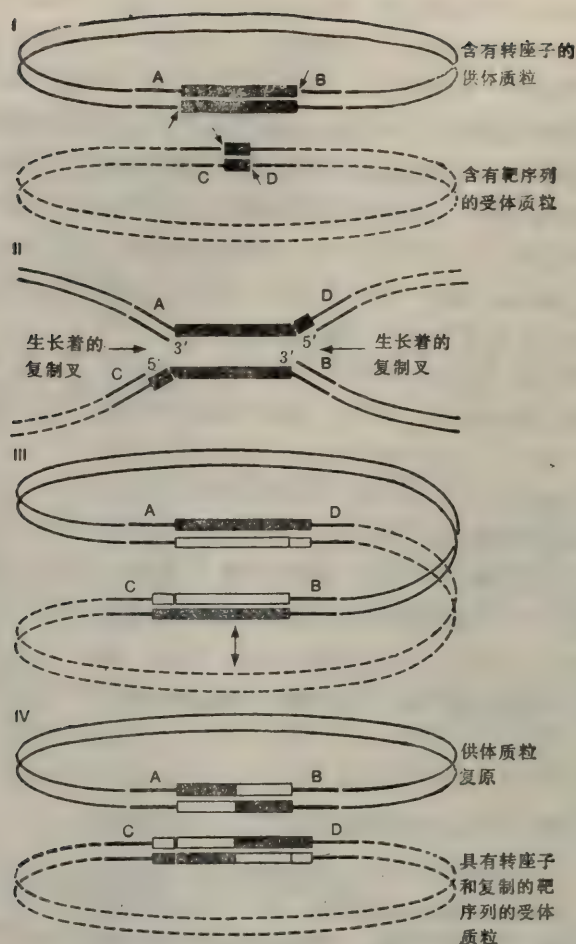


图 19-16 转座作用的一种模型 (由 J. A. Shapiro 提出, 见参考文献 S. N. Cohen 及 J. A. Shapiro, 1980)。(I)按箭头所指处将 DNA 分子在一定位点上剪开形成缺刻。(II)链在缺刻之间分开,与非同源链进行连接,结果产生复制叉。(III)通过 DNA 合成(中空线),形成图中框线所示的结构。具有专一性位点的链在内解区(箭头)之间发生链的交换。或者也有可能转座子的任何地方发生依赖于同源性的链的交换。(IV)质粒分开。大写字母表示碱基序列。

I. 在靶序列上打开两个缺刻(通过转座酶, transposase?), 这两个缺刻是交错的,因此当靶序列被打断时,产生两个互补的单链。

II. 在转座子的不同链的、相对的两个末端上打开两个相对的单链缺刻; 这样产生两个自由的末端。通过每一个自由的末端连接到靶位点序列上交错切割所产生的、突出的单链上,产生两个复制叉。

III. 以靶序列的突出末端为模板, 合成一条互补链, 开始其复制过程。当复制完成以后,形成含有两个拷贝的转座子的共整合物。

IV. 最后, 在两个拷贝的转座子之间发生双链交换过程。这个交换可能发生在末端重复的地方, 而在 Tn 3 中, 很可能发生在内解区处。交换的结果是将共整合物分开成供体单位与受体单位, 其中各含有一个拷贝的转座子。在受体中, 转座子两翼有着正向重复的靶序列; 其长度就是交错切口之间的碱基序列的长度(即相当的碱基对数目)。



下面我们将讨论另一种称之为不对称的模型 (asymmetric model)，这个模型也已普遍地为人们所接受，它是几种模型的综合物。与上面的那种模型有两点重要的区别：(1) 在转座子两个末端的切割作用不是同时进行的，而是在插入过程的不同阶段进行的；(2) 共整合物不是一个中间物，而只是两种不同产物中的一种，另一种只是简单的插入，而且，共整合可能发生在这个过程的最末时刻。图 19-17 给出了这个模型的图示。它的

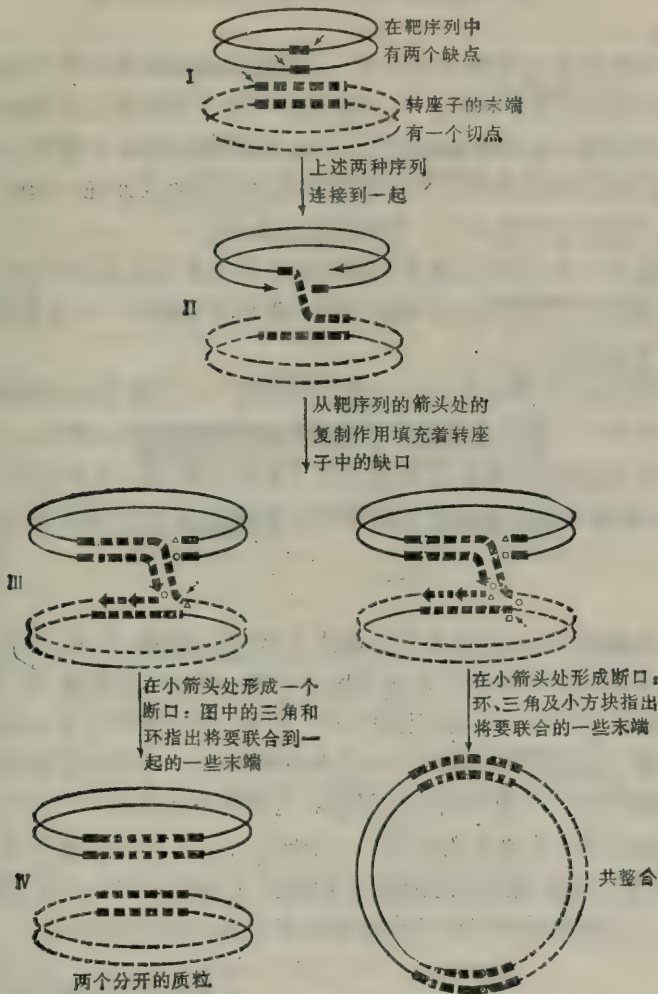


图 19-17 转座作用的不对称模型。这个模型以简单的插入或形成共整合物为结束（见图中第 IV 行，分别列于左右的两个小图）。第 III 行图中所示的两个结构是相同的；画出两个图只是为了清楚地说明为了得到第 IV 行图示的结构所需要进行的切割与连接。

步骤如下(文中的编号与图中的相同)。

- I. 在靶序列上产生两个切口，但是在转座子上只在左端有一个切口。
- II. 靶序列的一个末端与转座子的自由末端连接到一起。
- III. 然后进行复制，形成如图所示的中间物。注意，复制终止于转座子的末端，推测可能某些碱基序列是它的终止信号。
- IV. 复制作用终止之后，可能发生两种切割与连接，产生带有正常复制作用的转座子

简单插入的质粒（第 IV 行左图）或共整合物（第 IV 行右图）。

目前，只有很少的证据支持这个模型，这个模型应当被看作是不对称模型的一种雏型。

现在人们正在大力进行研究，以便对众多的模型（主要有两种类型）作出正确判断。

## 转座子介导的其他遗传学现象

转座子参与许多遗传现象，例如这一节将要讲到的基因重排与质粒-染色体整合等。

基因重排在进化中非常重要，它可能是由转座子介导的。例如转座子可能负责将一些原来在染色体上相距甚远的基因拉在一起，组成现在的被构建到一起的协调地控制着的一些操纵子。也有可能把两段分离着的碱基序列转变成连在一起的 DNA。这样的大踏步前进的进化过程就可能产生一些新的蛋白质分子。

在细胞繁殖的一个世代中，基因重排可能也是非常重要的。例如，可转座因子可以被看作是一个开关；在一个特殊的位点发生插入作用或是倒转一个含有启动子的序列，就可以使基因打开或者关闭。

在下面的讨论中，值得注意的一点是下面的每一个现象与过程都是由转座子介导的，但并不发生如图 19-16（第 IV 行）中所示的那种最后的解体步骤。然而，不能以为这些事件是不正常的，因为所有缺乏解体体系（例如，IS 因子）的转座子都发生基因重排等现象；一些虽然能够解体但是其解体频率很低的转座子，也都能发生这些现象。

### 复制子融合

图 19-16 的检测指出，如果不发生解体步骤（图中步骤 IV），那么发生图中步骤 I 所示的转座子介导的两个环形分子的重组融合过程，就会形成步骤 III 所示的一个单环分子。这个新的 DNA 分子含有原来两个 DNA 分子的所有基因以及两个拷贝的转座子。但是要注意，在这个过程中，每一个拷贝的转座子都没有复制的两翼靶序列，代之的是一个转座子与一个靶序列互相邻接。这种类型的重组作用已在两个质粒之间观察到，两个质粒中的一个携带着原有的一个转座子，这个过程称作质粒融合（plasmid fusion）。它只是更为普遍的称之为复制子融合（replicon fusion）现象中的一种特例。复制子融合是指由转座子介导的两个完整的复制体系的融合。

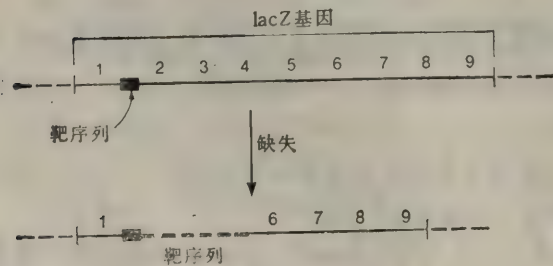


图 19-18 转座子造成的缺失的特点。2,3,4,5 片段都被转座子取代而缺失。

### 由转座子引起的缺失与倒转

当转座子出现在染色体或质粒中时，其邻近序列发生缺失的频率比无转座子时大 100—1 000 倍。在发生缺失时经常发现（1）转座子仍存在；（2）缺失是从转座子的末端开始的；（3）转座子只与一个靶序列拷贝相接（也就是转座子的两翼没有正向重复的序列）。图 19-18 给出了这些特点。这个现象在



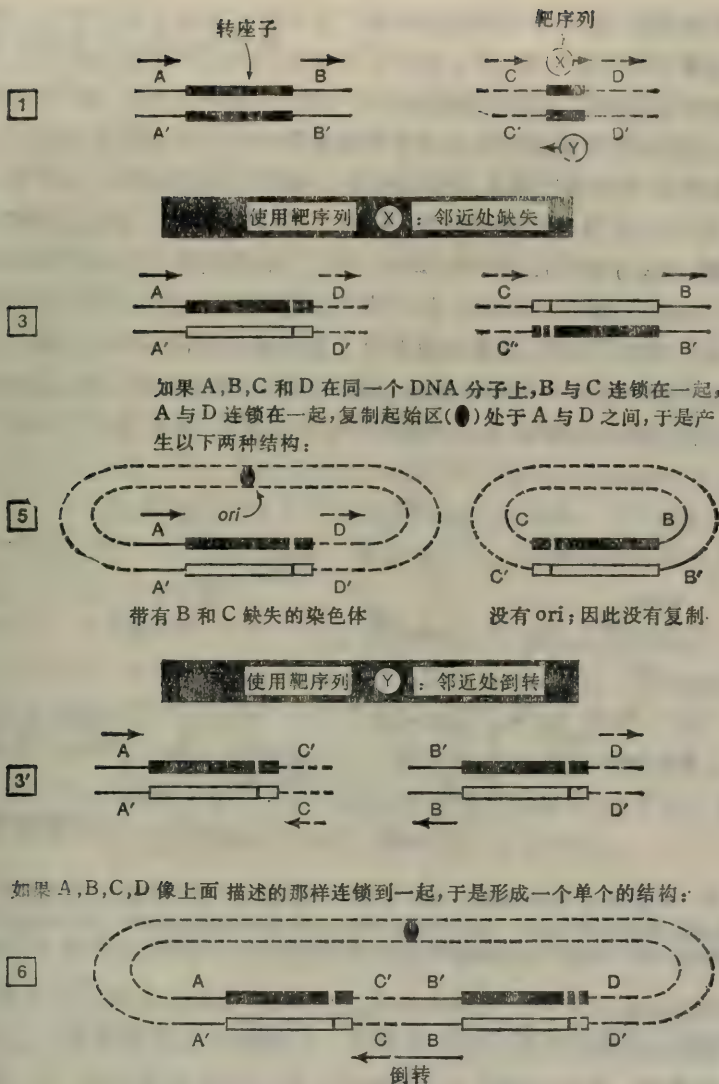


图 19-19 按图 19-16 中的 Shapiro 模型图解出缺失与倒转是怎样产生的。I 行所示的排列与图 19-16 中的相同。除了使用两种不同方向的靶序列之外, 3 与 3' 对应于图 19-16 中的 III。实心粗箭头表示转录方向, 其他符号与图 19-16 所用的一样。

19-19 中得到了简单的解释, 这个解释可用图 19-16 的模型表示。

在这个图中, 又重复了图 19-16 所示的那种排列, 也就是说, A、B 片段被转座子分割开, C 和 D 片段夹着靶序列, 与前面两个图中所示的步骤 I 的图及 1 片段一样。步骤 3 表示通过图 19-16 中的步骤 I 之后到步骤 III 时, 围绕着两个转座子的排列已经完成。如果步骤 1 中所用每一个 DNA 片段都是环状的, 那么步骤 3 中的两个片段就会融合成一个环。这些在前面的章节中已经作了解释。然而, 如果转座子的两个片段都来自一个单个 DNA 分子 (例如, 同一个染色体), 转座作用将产生步骤 5 所示的两个分子。注意, 在这个步骤中只有一个分子含有复制起始点, 而且这个分子也丢失了含有 B 和 C 的片段; 也就是说这个分子缺失了 B 和 C。如果含有 B、C 片段的较小的环不含有复制起始点, 就不能进行复制。因此, 如果这整个的过程在一个细菌中发生, 那么当细胞进行分裂时, 这个

较小的环就不能增多,这样子代细胞中就会产生只含有发生了缺失的染色体。注意,有缺失的部分紧挨着转座子的一侧,并且包含了转座子与靶序列之间的所有物质。

图 19-19 的下面部分表示倒转是如何发生的,在这种情况下,基因初始时的排列与产生缺失时是一样的,但其靶序列的方向与转座子方向相反(也就是说,按图 19-16 步骤 I 的相反方向连接)。这相当于转座子按相反的方向插入,然后转座过程产生步骤 3' 中的结构。注意, A 和 C 的排列与开始时的排列相比,其方向发生了倒转, C 和 D 也是同样。所以,如果起始的一些片段来自同一个 DNA 分子,结果就会产生步骤 6 的排列方式(在这种情况下,并没有发生缺失 DNA),但是处于转座子的一侧与靶序列之间的片段,与另一个片段相比较就发生了倒转,也就是说发生了基因倒位 (genetic inversion)。

虽然绝大多数转座子所介导的倒转过程都是按这种方式进行的,即没有解体过程;但其他的一些倒转过程也有可能通过解体或很有可能通过 *recA* 蛋白介导的重组来完成,图

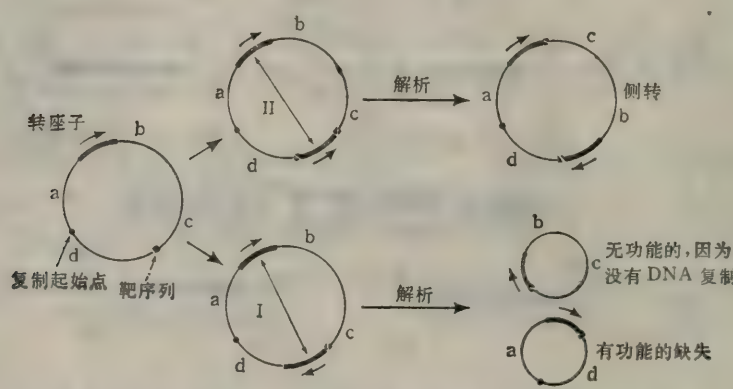


图 19-20 产生缺失与倒转的模型。转座子 DNA 按 I 或 II 两种方向插入靶序列中。图中的环可以是质粒,也可以是染色体。通过具有专一性位点的链在位点处的交换(双箭头指出之处),或者通过同源序列的交换(双向箭头所示)的解体过程,从 I 产生出缺失,从 II 产生出倒位。

19-20 的图解对此给出了一个简单的机理。这个转座过程发生在一个含有转座子的 DNA 分子的靶序列之内(即这个分子既有转座子,又有靶序列,在图中最左边的那个环上用小的方块表示靶序列)。所发生的插入过程既可以是正向插入(图中的 I),也可以是反向插入(图中的 II)。如果某个转座子能够解体,那么在它的两个转座子之间就会发生专一性位点的交换,也可能会发生由转座子基因或 *recA* 蛋白介导的依赖于同源性的交换(如果细胞是 *recA*<sup>+</sup>)。如果转座子是以正向排列,交换后产生两个环。如图 19-19 所示,因为染色体或质粒一般只含一个复制起始点,所以只有拥有复制起始点的片段才能存在下去;但是这个片段与原来的相比,已经丢失了一段 DNA 片段,因此它是缺失突变体。如果转座子是反向排列的,那么就会产生出像图中右边部分那样的倒转序列。

注意图 19-19 与图 19-20 所示的最终结果是相同的。

## 利用同一转座子两个拷贝之间同源性的过程

一些已知的专一性交换的过程是由 *recA* 介导的、在多个转座子拷贝之间进行同源重组的结果。下面讨论其中的两个: Hfr 细胞的形成和基因的扩增作用。



### 在 Hfr 形成过程中 IS 序列的作用

通过 F 因子的整合会产生 Hfr 细胞。几个证据（主要是确定 F<sup>+</sup> 质粒的结构）表明整合了的 F 序列的两翼总是有 IS 序列的两个拷贝（它们是正向重复的），这种 IS 序列是存在于 F 质粒中的序列；另外，在染色体中也发现了 F 质粒中所带的 IS 序列，这两个事实说明转座子参与了 Hfr 的形成。

含有 F 质粒的细胞可以通过两种方式形成 Hfr 细胞。一个方法是通过两个相同的 IS 序列发生同源重组而形成 Hfr 细胞；这两个 IS 序列一个在染色体上，另一个在 F 质粒上。如果这个重组事件是物理性的交互过程，那么按照这个机理就会在转座子的两翼产生两个 IS 序列，并且可能利用细菌的 Rec 体系、转座子基因产物或 F 质粒中的可以想见的某些基因产物形成 Hfr 细胞。另一种形成 Hfr 细胞的途径则是通过 F 质粒中 IS 序列所介导的共整合途径以及通过染色体中靶序列的复制（例如复制子融合）。很重要的一点是应当认识到这两种机理的结果是同样的。也就是说，它们的结果都是正向重复的两个拷贝的 IS 序列夹着 F 因子。Hfr 细胞通过以上两个途径而形成 Hfr 细胞是完全可能的，其原因将在下面接着叙述。

F 整合存在着许多个位点。但在一些被称为热点（hot spots）的位点上，比起其他的位点（称之为次位点 minor sites）更容易发生整合作用。现在已经测出包括热点在内的几段染色体片段的碱基序列，在每一个片段中都发现有 IS 序列。更进一步，大多数 Hfr 细胞曾经是  $rec^+F^+$  细胞，因此似乎很可能这些 Hfr 细胞株是由  $recA$  蛋白所介导的同源重组作用产生的。发生在次位点的整合作用频率低得多，而且常常发生在  $recA^-$  细胞中；最终的 Hfr 细胞株可能是通过共整合途径而产生出来的。

### 细菌中的基因扩增

$recA$  蛋白所介导的、转座子的多个拷贝之间的重组作用被人们用来解释另一个现象：细菌中的基因扩增。

一些 R 质粒（但不是所有的）有这样一种性质：如果把含有这种质粒的细胞放在含模式 I

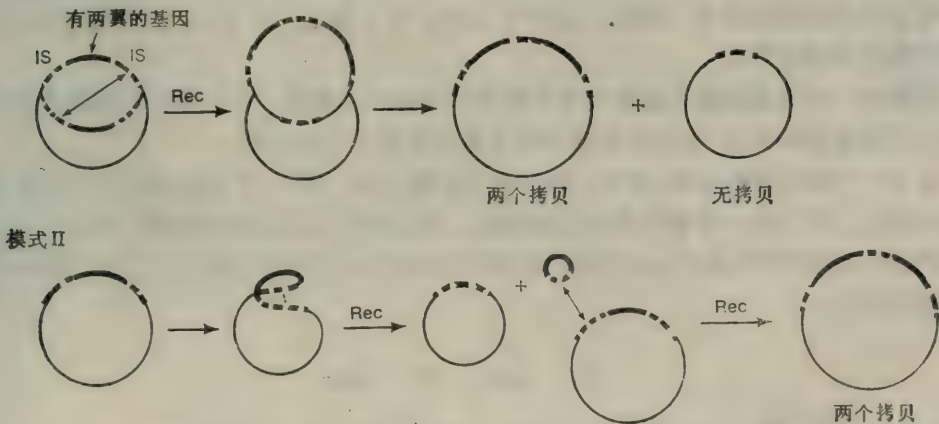


图 19-21 基因扩增的两个模型。双箭头表示 Rec 介导的 IS 序列之间发生的重组作用。在第 I 个模型中，重组发生在复制期间；在第 II 个模型中，有两次 Rec 介导的重组作用。第二次重组发生在已切除的小环与原初质粒的另一个拷贝之间。

有低于其最大耐受量的抗生素的培养基上进行培养,经过很多世代以后,这些细胞就对逐渐提高浓度的抗生素有抵抗力。这种抵抗力是因为 R 质粒中的抗药性基因逐渐增加所致。也就是说,当抗药性基因不断地串联复制时,R 质粒的大小就不断地增大。这个过程中要求有一个有活性的细菌 *rec* 体系,说明这个复制作用不是一个转座过程。基于质粒中的抗药性基因的两翼有转座子的事实,图 19-21 中给出了基因扩增的两种模型(其实还有其他的模型,只是这两个特别简单)。注意,这两个模型有两个基本要求,即首先是待扩增的基因的两翼有相同的序列,第二是 *Rec* 体系能够有效地使同源区进行配对。但是现在还没有支持这些模型的事实。

## 真核细胞中的转座子

首先,证明生物中存在着转座因子的证据不是来自细菌,而是来自 Barbara McClintock 对玉米的巧妙研究。她的发现指出,玉米中一些遗传位点的活性受到一种控制因子的影响,而且只有当这个因子靠近这些位点时,它才能抑制这些位点的活性;这些位点的基因活性水平,只有当这种控制因子进入或离开它的抑制位置时才发生变化。最近,物理图谱的数据说明玉米中的这个控制因子就是转座子。

在啤酒酵母 (*Saccharomyces cerevisiae*) 和黑色果蝇 (*Drosophila melanogaster*) 中也发现了转座子,还分离到了酵母转座子中的一个  $\delta$ 。 $\delta$  因子含有 338 碱基对,在每一个单倍体细胞中约含有 100 个拷贝。一个类转座子样的片段 Ty1 也已从酵母中分离出来。Ty1 大约有 5 100 碱基对,两翼有两个拷贝的  $\delta$ 。每一个单倍体细胞中约含有 35 个拷贝的 Ty1。Ty1 的插入能引起极性突变,转录的起始点靠近  $\delta$  的末端。在那些少数已知碱基序列的转座过程中, Ty1 插入时,它的两翼总伴随着有 5 碱基对的序列——这是大肠杆菌中的转座子信号。在果蝇中约有 3% 的 DNA 组成由约为 5 000 碱基对的各种倒转重复序列组成。其中大多数片段含有大约 50—200 碱基对。有一个叫做 *copia* 的转座子约含有 4 400 碱基对,它的两翼是 300 碱基对长的正向重复序列。*copia* 发生转座时影响某些基因的活性,推测可能是由于调节位点发生移动造成的。例如启动子以及其他一些蛋白质的结合位点(这些位点位于 *copia* 中)移动了。*copia* 的两翼也有一对 5 碱基对的正向重复序列。

果蝇的一些遗传现象可以用转座子的存在来解释。例如,有一些能抑制邻近的重组作用的突变是极性突变,而且已知这种突变是由于插入而造成的。

还有一个很重要的发现,就是一些(也可能是全部)RNA 肿瘤病毒本身可能就是转座子;然而,目前只有一个事实支持这个观点,即几种整合后的肿瘤病毒的两翼有 600 碱基对的正向重复序列;此外,还有短的靶序列的复制品。这些将在第二十二章肿瘤病毒中进行深入讨论。

## 参 考 文 献

- Bukhari, A. I. 1981. "Models of DNA transposition." *Trends Biochem. Sci.*, 6, 56—60.  
Bukhari, A. I., J. A. Shapiro, and S. L. Adhya, 1977. *DNA Insertion Elements and Plasmids*. Cold Spring Harbor Laboratory.  
Calos, M. P., and J. H. Miller. 1980. "Transposable elements." *Cell*, 20, 579—595.



- Campbell, A. 1981. "Evolutionary significance of accessory DNA elements in bacteria." *Ann. Rev. Microbiol.*, 35, 55—84.
- Campbell, A., M. Benedeck, and L. Hefferman. 1979. "Viruses and inserting elements in chromosomal evolution." In A. S. Dion, ed. *Concepts of the Structure and Function of DNA, Chromatin, and Chromosomes*. Symposia Specialists.
- Cohen, S. N. 1976. "Transposable genetic elements and plasmid evolution." *Nature*, 263, 731—738.
- Cohen, S. N., and J. A. Shapiro. 1980. "Transposable genetic elements." *Scient. Amer.*, February. pp. 40—49.
- Green, M. M. 1980. "Transposable elements in *Drosophila* and other Diptera." *Ann. Rev. Genetics*, 14, 109—120.
- Kleckner, N. 1977. "Translocatable elements in procaryotes." *Cell*, 11, 11—23.
- Kleckner, N. 1981. "Transposable genetic elements." *Ann. Rev. Genetics*, 15, 341—404.
- McClintock, B. 1965. "The control of gene action in maize." *Brookhaven Symp. Biol.*, 18, 162--184.
- Nevers, P., and H. Saedler, 1977. "Transposable genetic elements as agents of gene instability and chromosomal rearrangements." *Nature*, 268, 109—115.
- Starlinger, P. 1977. "DNA rearrangements in procaryotes." *Ann. Rev. Genetics*, 11, 103—126.
- Starlinger, P., and H. Saedler. 1976. "IS elements in microorganisms." *Current Topics Microbiol. Immunol.*, 75, 111—152.

## 第二十章 重组 DNA 及遗传工程

在过去的 40 年中,一些新技术的发展使人们对分子生物学的理解向前迈进了一大步。例如,X 射线衍射使研究者能对大分子结构进行明细的观察,氯化铯梯度离心技术则为认识 DNA 复制的机理敞开了大门。最近的新进展则是体外连接 DNA 分子的技术,即重组 DNA 技术。这项基本技术和它的许多发展已经彻底地变革了人们对真核生物的研究,并且已经对一些至今还被人们认为是几乎无法大量取得的特殊蛋白质提供了大量获得它们的办法。

本书到此,已经着重地介绍了分子生物学中的各种重要现象、基本概念及生物学机理,而对技术方面讨论得较少。但是,重组 DNA 技术或遗传工程在最近这些年来则是一个很重要的研究方面,确实应当给以充分的注意。

### DNA 分子的连接

重组 DNA 技术的基本操作包括两个步骤:(1)将 DNA 片段(这个片段因某些原因而使人们感到兴趣),连到一个可以复制的 DNA 分子上;(2)提供一个可让以上 DNA 连接物进行繁殖的环境(见图 20-1)。当做完这部分工作后,就说供体片段上的基因被克隆了,而携带分子就是载体或克隆运载体。

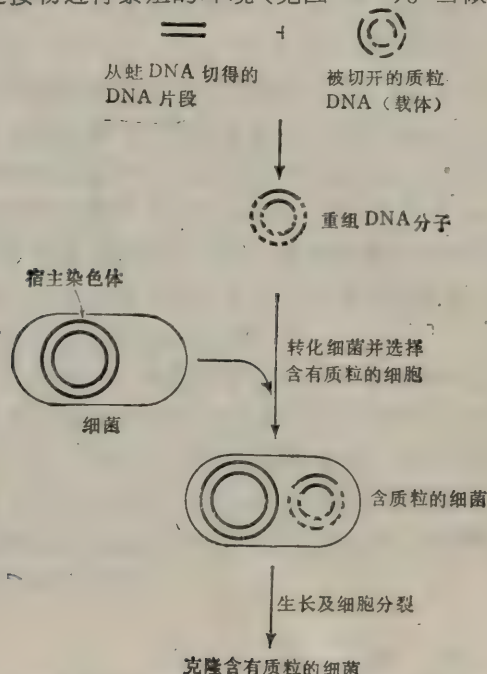


图 20-1 一个克隆的实例。蛙 DNA 的一个片段被结合到切割过的质粒载体上。杂交载体转化一个细菌。进一步,通过质粒的复制蛙卵 DNA 被携带到所有的子代细菌中。

在这个部分将要介绍 DNA 片段与载体连接的操作步骤和许多类型的可以使用的载体。

### 载体

为了成为一个有用的载体,载体必须具备三种性质:

1. 它必须能复制。
2. 必须有可以检测其存在的某些方法。
3. 必须有办法将载体 DNA 引入细胞。

使用得最普遍三类载体是质粒、大肠杆菌的  $\lambda$  噬菌体和病毒。因为这三种中的每一种的 DNA 都符合上面提到的三种性质。例如,如果把细菌细胞分子悬浮于冷的  $\text{CaCl}_2$  溶液中,那么 DNA 分子可以透入到许多种细菌的细



胞中。如果这种 DNA 分子是质粒 DNA，某些时候质粒会稳定地存在于细菌之中。在大部分的实验方案中，每个细胞中大约可进入 30 个 DNA 分子，在大约 0.1% 的细胞内质粒可稳定地建立于其中。含质粒的细胞可以很容易地从一般细胞中分离出来，有关分离的操作步骤将在本章后面一些部分介绍。当使用  $\lambda$  噬菌体的 DNA 时， $10^4$  个受感染的细胞中的一个细胞会产生子代噬菌体。对于酵母 (yeast, 或 *Saccharomyces*) 及其质粒，也已经发展出转化的操作步骤。一般，动物病毒很容易透入到动物细胞中，加入  $\text{Ca}_3(\text{PO}_4)_2$  稀溶液将使进入细胞的 DNA 量和穿透的速度会有一定的增加。以这种方式感染的动物细胞只有一小部分产生子代病毒。接着，或是让具有质粒的细胞生长，或是让噬菌体或病毒感染细胞。作为这些过程 (对于质粒 DNA 来说，叫做转化作用；对于噬菌体 DNA 或病毒 DNA 来说，叫做转染作用) 的结果，是重组 DNA 分子得以维持并繁殖。

普遍使用的只是少数的载体；其中的某一些和它们的性质列于表 20-1 中。

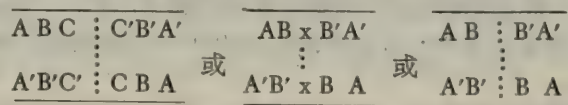
表 20-1 现在使用的某些克隆载体

命 名	基因型或性质
质粒	
pSC 101	tet <sup>+</sup>
Col E1	imm E1
pMB 9	tet <sup>+</sup> immE1
pBR 322	tet <sup>+</sup> amp <sup>+</sup>
pBR 325	tet <sup>+</sup> amp <sup>+</sup> cam <sup>+</sup>
Yeast 2 $\mu$ m	None
噬菌体	
$\lambda$ gt4- $\lambda$ B	可以溶源化；是温度可诱导的
$\lambda$ -Charon	lacZ <sup>+</sup>
$\lambda\Delta$ z1	插入作用发生于 lacZ 基因内
M13mp2	含有 lacP, lacO, lacZ 基因 单链 DNA；在序列分析中很有用
病毒	
SV40	病毒感染动物细胞。可以加进去的 DNA 的大小不能大于 $3 \times 10^6$ 分子量

缩写: tet, 四环素; imm, 免疫性; amp, 氨苄青霉素; cam, 氯霉素; lacP, 乳糖操纵子的启动子; lacO, 乳糖操纵子的操纵基因; lacZ, 编码  $\beta$ -半乳糖苷酶的基因。对于抗生素基因来说, + 表示抗性。

限制性内切酶

限制性内切酶是可以识别 DNA 分子中专一性序列的一类 DNA 内切酶,它分别在 DNA 双链分子的两条链上打下一个切口,产生 3'-OH 及 5' 末端。大约已从 250 种不同的微生物中分离纯化了 400 种左右的不同的限制性内切酶。在这么多的酶中 几乎只有少数的一些酶能够识别具有旋转对称性的序列,即其识别位点有如下形式的序列:



在上面的这些形式中,大写字母表示碱基,一撇表示互补的碱基, x 是任何碱基,虚线

代表对称轴。现在已经知道有多于 6 个碱基的限制性内切酶识别序列，但是还没有见到少于 4 个碱基的识别序列。

存在着两种不同类型的限制性内切酶。类型 I 酶可以识别专一性的序列，但其切割作用位点则在另外的地方，类型 II 酶的切割作用位点就在于识别的专一性序列之中\*。在本章我们仅仅考虑类型 II 酶。所有的限制性内切酶，都造成两个单链缺口，每条链上一个。这些缺口可能有两种不同的排列方式：

1. 形成平整末端 (flush 或 blunt ends): 两个缺口都在对称中心处。
  2. 形成粘性末端 (cohesive ends): 两个缺口分别对称地分布在对称轴的两侧。
- 这些排列及酶作用的后果见图 20-2。

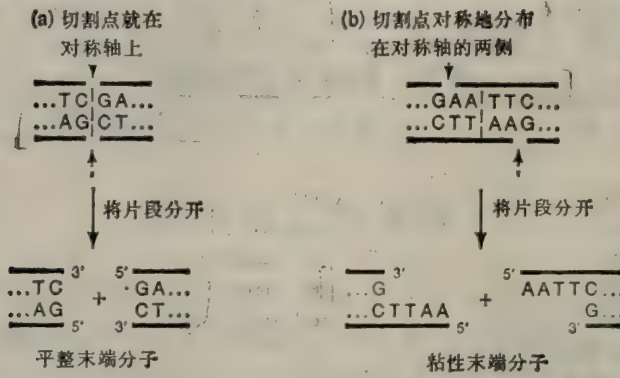


图 20-2 限制性内切酶造成的两种切割类型。箭头表示切割位点，虚线表示对称序列的中心。

表 20-2 列出了 12 种比较有用的限制性内切酶的识别序列及切割位点，其中 9 种产生粘性末端，3 种产生平整末端。

因为每一种限制性内切酶只能识别一种独特的序列，因此这些限制性内切酶都有一个重要的特点，那就是对于从某一种有机体中分离出来的 DNA 来说，一般其作用位点的数目是较少的。例如，典型的细菌 DNA 分子大约有  $3 \times 10^6$  碱基对，经限制性内切酶作用后，被切成几百个片段；更小一些的 DNA 分子如噬菌体 DNA 或质粒 DNA 受限制性内切酶作用，酶的切点可能少于 10 个（更常见的情况，只有 1 个或 2 个，有时连一个也没有）。我们将很快可以看到那些对于特定的限制性内切酶只提供一个作用位点的质粒，它们特别值得选来用作载体。

因为前面已经提到过限制性内切酶作用的专一性，因此，一种特定的限制性内切酶对某种特殊的 DNA 作用，可产生一族独特的 DNA 片段。而另一种限制性内切酶作用于同样的这种 DNA 分子上，将会产生另外一族独特的 DNA 片段（但也有些酶——同功异源酶 (isoschizomers)，它们彼此有相同的专一性，产生同样一族 DNA 片段）。图 20-3 (a) 表示了大肠杆菌  $\lambda$  噬菌体的 DNA 被限制性内切酶 EcoRI 及 Bam HI 切割的位点。被单个酶作用产生的片段族通常可以通过消化过的 DNA 样品进行凝胶电泳而检出（见

\* 近来，又在自然界中认出了类型 III 酶，它的活性很像类型 I 酶，但它本身是由不同的基因所编码的多个亚基构成。



表 20-2 某些限制性内切酶及它们的切割位点

微生物	酶的名称	靶序列及切割位点
产生粘性末端		
<i>E. coli</i>	EcoRI	$\begin{array}{c} \downarrow \\ \text{GAA} \text{ }   \text{ } \text{TTC} \\ \text{CTT} \text{ }   \text{ } \text{AAG} \\ \uparrow \end{array}$
<i>Bacillus amyloliquefaciens</i> H	BamHI	$\begin{array}{c} \downarrow \\ \text{GGA} \text{ }   \text{ } \text{TCC} \\ \text{CCT} \text{ }   \text{ } \text{AGG} \\ \uparrow \end{array}$
<i>B. globigii</i>	BglI	$\begin{array}{c} \downarrow \\ \text{AGA} \text{ }   \text{ } \text{TCT} \\ \text{TCT} \text{ }   \text{ } \text{AGA} \\ \uparrow \end{array}$
<i>Haemophilus aegyptius</i>	Hae II	$\begin{array}{c} \downarrow \quad \uparrow \\ \text{PuGC} \text{ }   \text{ } \text{GC Py} \\ \text{PyCG} \text{ }   \text{ } \text{CGPu} \\ \downarrow \quad \uparrow \end{array}$
<i>Haemophilus influenza</i>	Hind III	$\begin{array}{c} \downarrow \quad \uparrow \\ \text{AAG} \text{ }   \text{ } \text{CTT} \\ \text{TTC} \text{ }   \text{ } \text{GAA} \\ \downarrow \quad \uparrow \end{array}$
<i>Providencia stuartii</i>	Pst I	$\begin{array}{c} \downarrow \quad \uparrow \\ \text{CTG} \text{ }   \text{ } \text{CAG} \\ \text{GAC} \text{ }   \text{ } \text{GTC} \\ \uparrow \end{array}$
<i>Streptococcus albus</i> G	Sal I	$\begin{array}{c} \downarrow \quad \uparrow \\ \text{GTC} \text{ }   \text{ } \text{GAC} \\ \text{CAG} \text{ }   \text{ } \text{CTG} \\ \downarrow \quad \uparrow \end{array}$
<i>Xanthomonas badrii</i>	XbaI	$\begin{array}{c} \downarrow \quad \uparrow \\ \text{TCT} \text{ }   \text{ } \text{AGA} \\ \text{AGA} \text{ }   \text{ } \text{TCT} \\ \uparrow \end{array}$
<i>Thermus aquaticus</i>	Taq	$\begin{array}{c} \downarrow \\ \text{TC} \text{ }   \text{ } \text{GA} \\ \text{AG} \text{ }   \text{ } \text{TC} \\ \uparrow \end{array}$
产生平整末端		
<i>Brevibacterium albidum</i>	Ball	$\begin{array}{c} \downarrow \quad \uparrow \\ \text{TGG} \text{ }   \text{ } \text{CCA} \\ \text{ACC} \text{ }   \text{ } \text{GGT} \\ \uparrow \end{array}$
<i>Haemophilus aegyptius</i>	Hae I	$\begin{array}{c} \downarrow \quad \uparrow \\ \text{(A) GGCC (T)} \\ \text{(T) CC GG (A)} \\ \downarrow \quad \uparrow \end{array}$
<i>Serratia marcescens</i>	Sma I	$\begin{array}{c} \downarrow \quad \uparrow \\ \text{CCC} \text{ }   \text{ } \text{GGG} \\ \text{GGG} \text{ }   \text{ } \text{CCC} \\ \uparrow \end{array}$

注意：垂直的虚线表示每个序列中的旋转对称轴。箭头表示切割的位点。Taq I 酶产生含有两个核苷酸的粘性末端片段，而其他的酶产生的粘性末端片段都具有 4 个核苷酸。Hae I 酶识别 GGCC 序列，而其邻近的碱基对 A·T 或 T·A 也应对称地留在 GGCC 两侧。Pu 及 Py 分别表示嘌呤或嘧啶。

图 20-3 (b))。DNA 片段泳动的速度与分子量大小相关，这些片段的分子量可以参照同

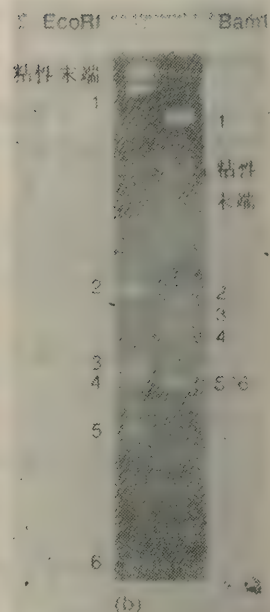
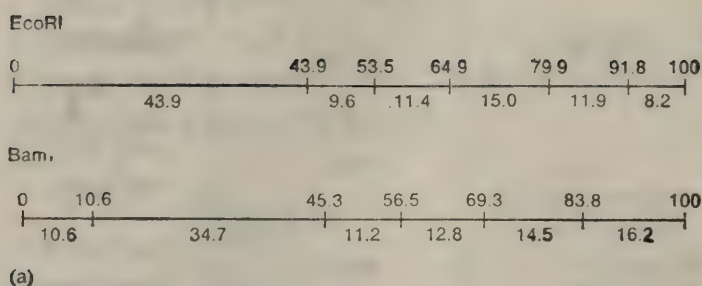


图 20-3 (a)  $\lambda$  DNA 对于 EcoRI 酶及 BamHI 酶的限制性内切酶图谱。垂直的棒表示酶的切割位点。上排数字表示从  $\lambda$  DNA 的 A 基因起到每个切点所占  $\lambda$  全长的百分率。 $\lambda$  全长是指测量从  $\lambda$  DNA 的 A 基因起到  $\lambda$  DNA 右侧末端止所得到的全部距离。下排数字是每一个片段的长度，长度也用总  $\lambda$  长度的百分数来表示。(b) EcoRI 酶及 BamHI 酶消化  $\lambda$  DNA 后得到的凝胶电泳图谱。用粘性末端标志的条带表示由  $\lambda$  DNA 的正常粘性末端连接两个末端片段所组成的分子所形成的条带。数字顺序表示从最大到最小的 6 个片段。BamHI 作用下 5、6 两个条带未能分开(承 Anderson 及 Lynn Enquist 提供)。

时泳动的已知分子量的片段所移动的距离而测出。

### 测定切割位点;限制性内切酶图谱

限制性内切酶在  $\lambda$  DNA 中的切割位置，即限制性内切酶图谱可用下面介绍的方法测定出来(图 20-4)。

1. 比较酶切线性 DNA 及酶切环式 DNA 的电泳泳动图谱。通过比较可以看出，环式 DNA 比线性 DNA 少两条相应于末端片段的条带，多具有一个由末端连接而形成的新条带。因此，认为处于线性 DNA 分子两个末端的片段是在环式 DNA 消化样品中不见了片段。再比较完全消化及部分消化样品的电泳结果。在一个典型的消化即完全消化反应中，并不一定每一个分子都被限制性内切酶切割，有时能看到未被切割的分子形成的极为细弱的条带。如果缩短酶反应的时间，就可以增强此细弱的条带，这种酶反应过程被称之为部分消化 (partial digest)。在这种部分消化反应之后，再进行电泳，除去极为细弱的条带外，将泳动图上其他条带与完全消化样品的泳动图作比较，部分消化样品中的这些条带比完全消化样品中相应的条带的红色荧光的强度弱得多。又发现增强条带的分子量必定是两个减弱条带所含有的片段的分子量之和。这表明减弱条带是原初的未切割分子，它由邻近的、靠在一起的两个片段组成。如果切割点的数量不很多(例如 6 个)，采用此法可以去弄清楚一些确定片段的排列次序。



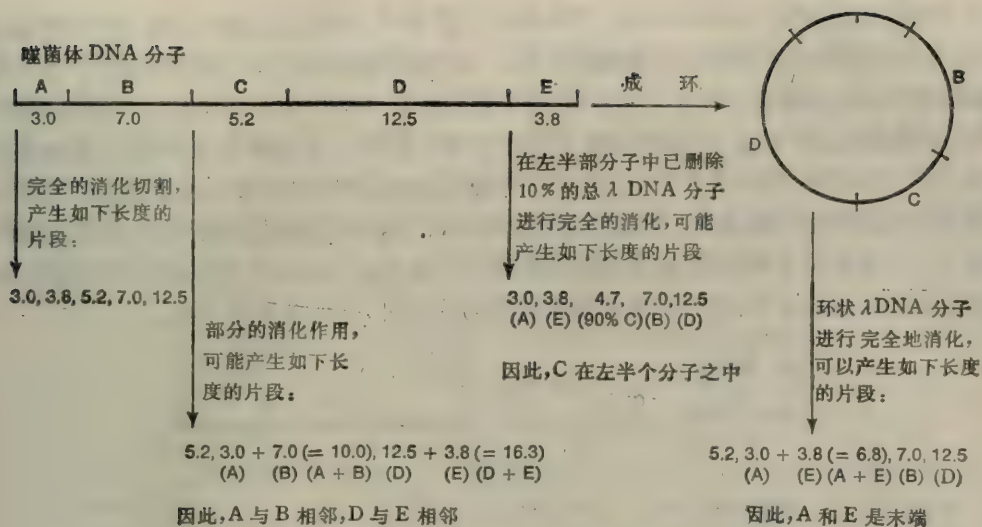


图 20-4 测定噬菌体 DNA 分子中限制性内切酶切点位置的可能图式。

2. 也还可以利用测定遗传删除突变体而鉴定出一个 DNA 片段在一个完整的 DNA 分子中应有的位置。如果某删除作用并不消灭一个限制性内切酶作用点, 那么就是在一个片段内去除一部分 DNA, 于是就会产生一个新的片段, 这个片段小于野生型中的相应的一个片段。新片段与野生型中的相应片段比较, 少去的正好是被删除的那一部分。反之, 如果删除的一段 DNA 中包含有限制性内切酶作用位点, 即去掉了一个限制性内切酶作用位点, 那么, 就会少去两个条带而出现一个新的条带。这个新的条带所代表的 DNA 的分子量相当于从上述两个下落不明的片段的分子量之和再减去被删除的那一部分 DNA 分子量。所叙述的这种类型的实验产生的切割图式 (限制性内切酶图谱) 可见图中。

3. 虽然前面介绍的方法通常是可行的, 但最普通的用于制作限制性内切酶图谱的方法是使用两种不同的限制性内切酶。其操作步骤为从同一种标定 DNA 中取三份样品, 对其中两份各加一种限制性内切酶, 对余下的那份加入以上两种酶 (双消化, double digest), 保温消化, 然后电泳, 再比较三份样品产生的三组片段。图 20-5 给出了一个例子, 表示可能产生的片段组。这个操作步骤的价值是可以得到两种酶共同作用的限制性内切酶图谱。我们假设, 通常也就是这种情况, 首先需要确定哪个片段是末端片段, 为达到这个目的, 或者用图 20-4 介绍的比较线性及环式 DNA 消化产物的方法, 或者在限制性内切酶作用之前, 用多核苷酸磷酸激酶对待测片段进行 5' 末端的  $^{32}\text{P}$  标记, 再于限制性内切酶作用后通过放射性来找出处于末端的片段。掌握了这项信息, 就能鉴定出与最末端的片段相邻的片段。首先, 让我们在样品 I 的消化产物中寻找邻近于片段 A 的片段。在双消化中, 那个在消化 II 中形成片段 F 的切点必然地要把消化 I 中邻近于片段 A 的那个片段分裂成两半, 从而产生出一个长度为  $2.3 - 1 = 1.3$  的小片段和另一个长度为  $x$  的小片段; 因此,  $1.3 + x$  必须等于片段 B、C 或 D 的长度。从双消化的结果去推断, 排除两端的两个末端片段, 即一边的 1.7 及另一边的 1.0, 在余下的各种片段中选择,  $x$  的仅有的可能值只能是 0.7。因此, 邻近 A 的片段必然是  $1.3 + 0.7 = 2$ , 这恰恰是片段 B 的长度, 所以 B 邻近于 A。现在我们来查找在消化 II 中邻近片段 F 的片段。B 的右端必定去分裂邻

近于 F 的片段。因为  $A + B = 3$ ,  $F = 2.3$ , 于是  $3 - 2.3 + y$  长度的片段的长度必然相等  
于消化 II 中邻近于 F 的那个非末端的片段,  $y$  仅有的可能值是 2.5。因此邻近于 F 的片段  
的长度为 3.2, 它必然是片段 G。我们再把注意力转到 DNA 分子的右端。片段 E (消化  
I 中) 的切点邻近于片段 J (消化 II 中) 的片段。因此, 在双消化 II 中有长度为 Z 的  
片段,  $5 - 1.7 + Z$  必须等于  $1.7 +$  在消化 II 中某长度的一个片段, Z 只可能是 5.4, 邻近于  
片段 J 的片段的长度就是 8.7 了, 这只能是片段 I。这个过程可以不断地继续下去, 最终  
能建立一个读者应当确认的、对于两种限制性内切酶都是完全消化的限制性内切酶图谱。  
这种方法在有的情况下, 即当存在着相同长度的片段时, 会更加复杂一些。

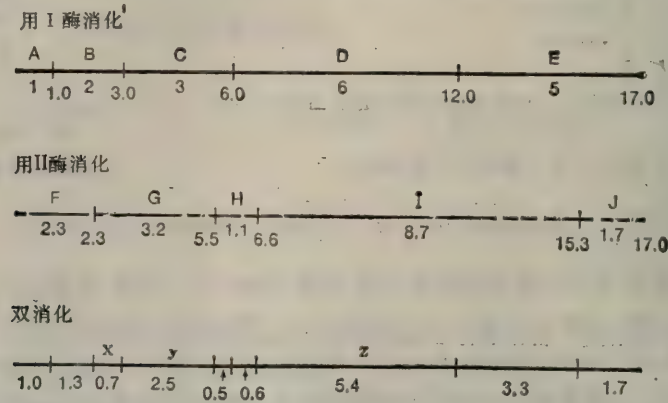


图 20-5 确定限制性内切酶图谱时三种限制性消化作用的用途。酶 I 产生消化 I, 酶 II 产生消化 II, 酶 I 及酶 II 共同产生双消化。虽然真正的次序要待数据进行分析后才知道, 但为了简化讨论过程, 先给片段一一编上次序。

### 由许多限制性内切酶产生的具有互补单链末端的片段

在研究限制性内切酶中最令人激动的事件之一是 1972 年 Janet Mertz 及 Ronald Davis 在斯坦福大学发现的事实: 如果把限制性内切酶产生的片段保存于高盐浓度的缓冲液中, 并置于适当的复性条件下, 许多限制性内切酶产生的片段会自动地环化。这个观察是证明某些限制性内切酶产生带有粘性末端的 DNA 片段的第一个依据。这个实验简示于图 20-6。

当然, 也如我们已经介绍过的, 并不是所有的限制性内切酶都产生带粘性末端的单链。因此, 有的只产生带有平整末端片段的分子, 这种平整端的片段不能环化。

### 具有粘性末端的限制性内切酶片段的连接

一种特定的限制性内切酶只识别一种简单的碱基序列。因此, 某特定的限制性内切酶作用于一种 DNA 分子产生的粘性末端, 相同于此酶作用于另一种不同的 DNA 分子所产生的粘性末端(因为两种 DNA 分子都有被此种酶所识别的序列)。如果某个片段的两个粘性末端有互补的碱基序列, 那么此片段本身就可以环化; 同样地, 两个片段也可以通过粘性末端的互补性而配对。因此来自两种有机体的 DNA(例如细菌的及蛙的)可以配对粘结(如图 20-7 所示)。进一步, 在碱基配对之后再进行铈封连接, 则两个片段可以永久性地连接到一起。



连接产生的 DNA 不总是有功能的 DNA 序列。例如,考虑一个被切成 A、B、C 及 D 四个片段的 DNA 分子。重新将这四个片段进行装配当然可以产生原初的分子,其序列假设为 A、B、C、D。但是因为 B 与 C 有同样的粘性末端,因此 ACBD 序列的分子就有可能以与 ABCD 分子同样的概率产生。在重装配过程中,也有可能产生出含有多个某一片段的分子,例如 ABBB 或 ACCD 分子。因此,只产生出 1/4 有功能的序列。类似地,如果 DNA 分子可以被切成五个片段 A、B、C、D 及 E,那么再行组装时, A(BCD)E、

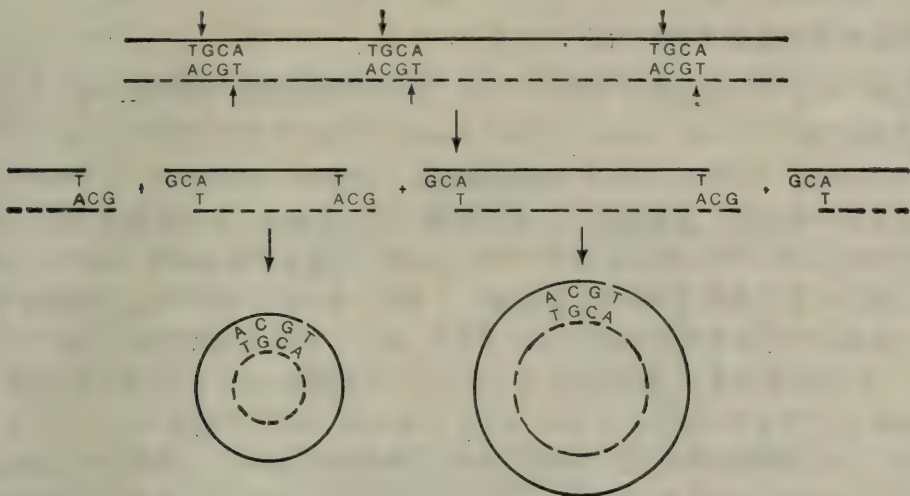


图 20-6 限制性内切酶可以环状化的一个实验。箭头表示切割位点。

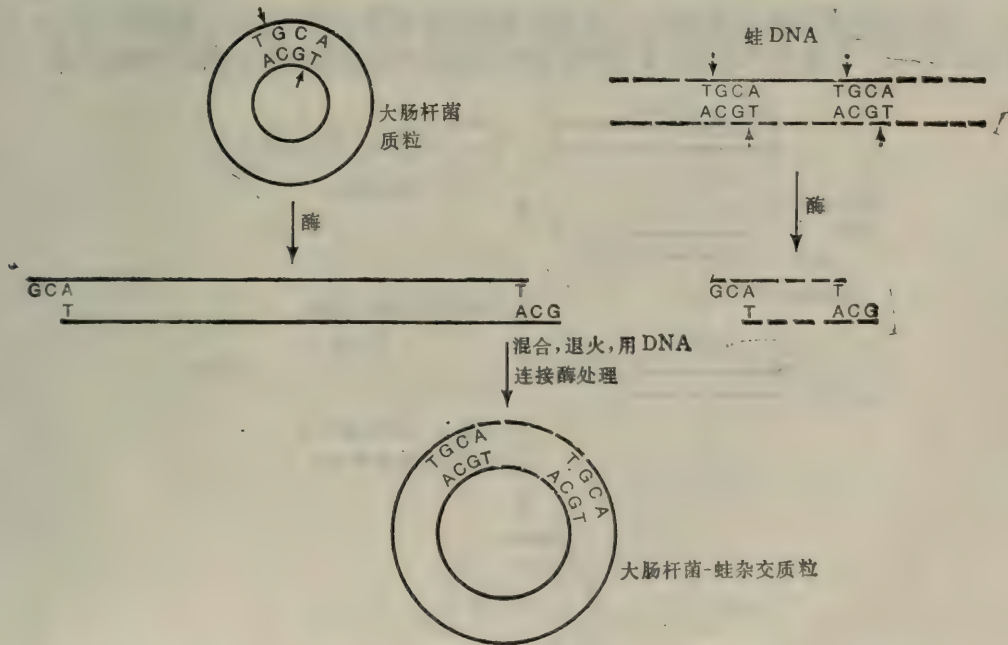


图 20-7 用限制性内切酶把来自不同有机体的片段组建成一个杂交 DNA 分子。这种种间杂交常常被称为嵌合体 (chimeras)。短箭头表示切割位点。

A(BDC)E、A(CBD)E、A(CDB)E、A(DBC)E 及 A(DCB)E 以及那些其中含有两个某片段或三个某片段的分子就都可能以相等的概率存在。在这种情况下,当把由两种不同 DNA 分子生成的片段进行再退火时,得到有功能的 DNA 分子的概率当然是很低的。因此,如果质粒 DNA 被切成四个片段,蛙 DNA 被切成 1 000 个片段,大部分组装的质粒将是杂乱拼凑并且无功能的。另一方面,如果用遗传学的方法选择出有功能的质粒,并且在 B、C 之间有可能有插入作用,那么得到含有任意的蛙 DNA 片段的杂交质粒的可能性非常高,但若要求插入一个特定的 DNA 片段,那么这种可能性将降低 1 000 倍。

### 通过外加均聚物连接 DNA 片段

重组 DNA 研究的领域开始于 1972 年,这恰好在限制性内切酶性质被人们了解之前。斯坦福大学的 Peter Lobban 及 Dale Kaiser 发展出连接任何两个 DNA 分子的一种一般性方法。他们的方法是采用末端转移酶 (terminal nucleotidyl transferase)。这种酶是从动物组织中分离得到的一种不寻常的 DNA 聚合酶,它把核苷酸(通过三磷酸核苷酸前体)加到 DNA 突出单链的 3'OH 上。酶作用时不需要模板链,因此这个过程很引人注意。为了形成突出的单链,应当先用 5' 端专一性的外切酶去移走少量的末端核苷酸。反应混合物中应有外切酶处理过的 DNA、dATP 以及末端转移酶,反应后在 DNA 的两个 3'-OH 末端形成 poly(dA) 尾(见图 20-8)。同样,如果供给 dTTP 为底物,那么 DNA 分子将会有 poly(dT) 尾。因此,如果多聚(dA)尾安装在一个 DNA 分子上,多聚(dT)尾安装到第二个 DNA 分子上,并令多聚(dA)与多聚(dT)退火,(如图所示),再利用 DNA 聚合酶 I 的缺口填补作用(gap-filling)使分子连接过程得以完成,最后可再用 DNA 连接酶将剩下的缺刻彻底封满。

这个方法叫做均聚尾连接法。此法对那些缺乏粘性末端又欲进行配对的 DNA 分子非常有用。这种分子可能是某些产生平整末端片段的限制性内切酶的作用产物,也

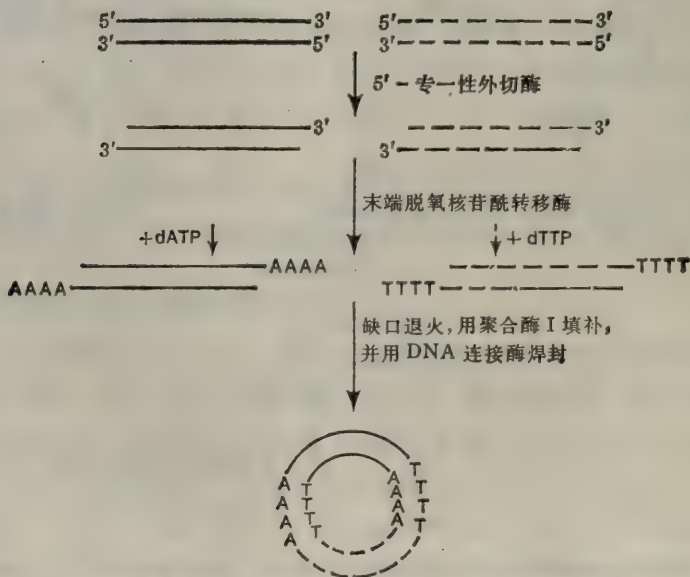


图 20-8 带有互补的均聚尾的两个 DNA 分子的连接。



可能是大的 DNA 分子经机械力作用而制备得到的产物,甚至可能是 cDNA 分子。cDNA 是在实验室中以 RNA 为模板,经逆转录酶作用而产生的 DNA 分子,它也是平整末端的分子,这种分子在重组 DNA 技术中有极端重要的作用。

平整末端的连接

大肠杆菌 T4 噬菌体编码一种 DNA 连接酶。受 T4 感染的大肠杆菌能产生大量的这种酶。它对双链 DNA 中的 3'-OH 末端及 5'-P 末端之间的缺刻可以焊封,因此是一种典型的连接酶。此外,它还有另一种特性,即可以把两个具有完全碱基配对的末端 DNA 分子连接到一起(见图 20-9)。这个反应是怎样发生的尚不清楚。但是对于带有平整末端的 DNA 分子的连接,使用起来效果很好,因此,它是有别于均聚物连接的又一个很有用处的方法。

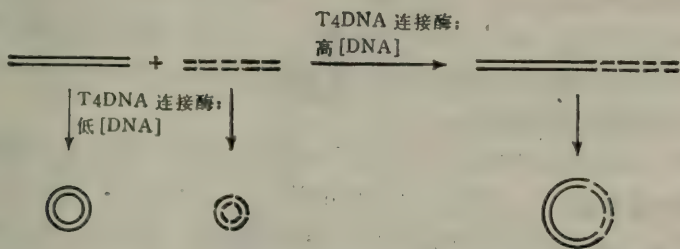


图 20-9 两个平整末端分子的反应。由图可见,在 DNA 浓度较低的情况下,优先发生分子内的环化作用。

用连接器连接

有时,一个带有平整末端的分子连到另一个具有粘性末端的分子上的做法非常有用,

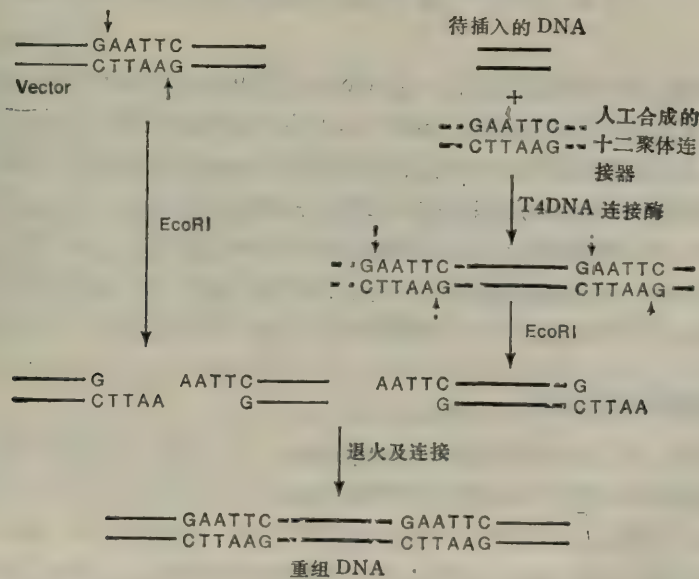


图 20-10 通过使用一个连接器而形成重组 DNA 分子。短箭头表示 EcoRI 酶的切割位点。

这里提到的具有粘性末端的片段是由某些产生单链末端的限制性内切酶所产生的。以上考虑是有可能实现的。如果两段短小的含有限制性内切酶位点的 DNA 片段(连接器, adaptor) 偶联到那个带平整末端的分子的两端上, 就可以达到以上目的。图 20-10 表示了实现上述目的的简明过程。在上述过程中, 平整末端分子得以插入到载体的 EcoRI 位点中。这种方法很有用处, 因为, 一旦需要回收插入的带平整末端的 DNA 片段时, 可以用连接器中有切割位点的那种限制性内切酶作用于重组分子上。但是, 如果采用的是均聚尾连接法将不能回收插入的片段。

### 利用流体动力的剪切力产生 DNA 片段

让我们先来考虑一个将被克隆到质粒中去的基因 G。最简单的克隆方法就是用同种的限制性内切酶作用于质粒 DNA 和含 G 的 DNA, 然后退火。但是有可能切割某质粒的各种限制性内切酶都在基因 G 上有切点, 从而使基因 G 失活。因此显然还必须要其他的切割 DNA 的方法。人们发现, 用机械法, 也就是用流体动力学的剪切力的方法代替限制性内切酶法可以使带有基因 G 的供体 DNA 断裂, 并可以消除专一性作用位点切割 DNA 所引起的基因失活作用。

如果使 DNA 溶液迅速地通过皮下注射用的针头或小孔, 或受到剧烈地搅拌, DNA 分子就被打断成分子量为  $2 \times 10^6$ — $20 \times 10^6$  的片段。某些 DNA 分子, 如基因 G, 它的典型分子量为  $10^6$ 。这些分子有可能被打断, 但是总的来说, 基因将保持其完整性。

剪切力作用下产生的分子通常都是平整末端的分子。虽然平整末端 DNA 分子可以直接地连接, 但是为了更有效地连接到那些已经经过限制性内切酶作用的质粒上, 加上一个合适的连接器接头可能更佳, 见图 20-10。

## 特定 DNA 分子加到载体中的插入作用

至此, 我们已经介绍了取得的一组 DNA 片段的各种方式(例如, 限制性内切酶消化或机械剪切), 以及该片段与载体一起退火, 产生大量的杂交载体的一些方法。如果要克隆一个特定的基因, 那么必定首先要从所有具有外来 DNA 的载体中分离出带有此基因的载体。虽然这种分离带有特定基因的杂交载体的做法在许多情况下是适用的。但是, 也确实还存在着许多其他的情况: 人们所感兴趣的克隆或者是太少, 或者是太难于检测。因此如果能提纯含有所感兴趣的基因的 DNA 片段人们还是宁愿先采用别的办法, 即进行杂交实验检测出所要的克隆。本部分将要介绍克隆一个特定 DNA 分子三类方法。

### cDNA 及逆转录酶的使用

某一特定基因是否已正确地插入到载体中, 可以很容易地通过观察基因的表达产物而证实。例如, lac 操纵子是否已正确地插进质粒, 可以通过它合成  $\beta$ -半乳糖苷酶的能力而被检出。但是去了解某一个真核基因是否已插入到细菌 DNA 载体则有可能会成为一个问题。因为, 一般地讲, 一个真核基因不能转录与翻译出一种有功能的蛋白质。但若采用本章所介绍的技术, 则检测插入的真核基因就是一个相当直接了当的工作了。



让我们假设,把兔的珠蛋白(血红蛋白的一个亚基)基因克隆到一个细菌质粒中\*。其过程为:用同一种限制性内切酶处理兔的细胞质 DNA 及大肠杆菌质粒 DNA,将片段混合,退火,连接,形成重组质粒,最后再用含有杂交质粒(即重组质粒)的 DNA 混合物去感染大肠杆菌。主要的问题是如何鉴别出那些含有携带着这种基因的质粒的细菌菌落。如果被克隆的是细菌的 gal 基因,我们就可以在适当的培养基上用 Gal<sup>-</sup> 宿主去选择 Gal<sup>+</sup> 菌落。然而,因为珠蛋白基因有内含子(见第十一章),而且所有的人都知道细菌缺少加工真核 mRNA 分子的酶,所以含珠蛋白基因的细菌不能合成珠蛋白。当然可采用其他的手段找出目的细胞,例如用纯的珠蛋白 mRNA 与细菌 DNA 杂交。但是因为 10<sup>9</sup> 个细胞中约仅有一个细胞含有带珠蛋白基因的杂交质粒,因此这个方法也许会因工作量太大而令人厌烦。本章的下一部分将为某些这类介绍另一类方法。

某些动物细胞,其中之一例是产生珠蛋白的网状细胞,只形成一种或少数几种蛋白质。在这些细胞中,已经加工过了的、专一性的 mRNA 分子构成细胞中总 mRNA 中的大部分。因此,通常获得的 mRNA 样品可以几乎只由一种 mRNA 组成,在这种情况下,只有珠蛋白 mRNA。所以若对那些最终的基因表达产物就是细胞蛋白质的主要成分的基因进行克隆时,这种类型细胞中的纯 mRNA 就可以作为起始物质,由它创建出仅仅只含有我们所感兴趣的基因的重组质粒。

从动物病毒中分离出的逆转录酶可以用 RNA 为模板合成出双链 DNA 分子(其机理,见第二十一章)。用这种方式制备的双链 DNA 分子叫做互补 DNA 或 cDNA。如果用作模板的 RNA 分子在从体内分离出来之前已被加工过(即内含子已被移走),那么与之相对应的 cDNA 就含有一个不会被打断的编码序列。因此,如果可以分离得到加工后的 mRNA,形成重组 DNA 分子的目的是在细菌中合成一种真核基因的产物,而且那么就可以选择加工后的 mRNA 所形成的 cDNA 为插入物质。cDNA 具有平整末端,一般可通过采用连接器方法把它连到载体上,如图 20-10 所示。

### 插入互补的人工合成基因

高等有机体,例如人,产生许多多肽激素,它们都有很大的医学价值,而且典型的组成是少于 20 个氨基酸残基的肽类。在许多情况下,多肽激素的基因或其 mRNA 都还未分离出来。目前,已经可能采用简单的技术去合成短小的、有一定序列的片段。又因为目前已经知道了大部分多肽激素的氨基酸序列。因此,可以由遗传密码推绎出特定基因的核苷酸序列。合成的基因的碱基序列不一定非要与天然基因的完全相同,因为(1)密码的冗余(redundancy)允许大幅度调换密码子,而并不改变信息的内容;(2)天然基因中可能包含有内含子,人工合成的则无。而且,为了生成一个有功能的人工合成基因,合成的 DNA 分子中必须包括起始密码和终止密码,碱基序列应以起始密码子开始,以终止密码子告终。一旦基因合成,就可以用已经介绍过的任何一种插入平整末端 DNA 片段的方法把合成基因连接到载体上。目前用这种方式合成的最大的 DNA 片段是有 514 个碱基对的人干扰素的基因,它足够用于合成出具有 170 个氨基酸残基的蛋白质。

在介绍生长激素释放抑制因子(somatostatin)基因的克隆时,我们将对合成基因进

\* 克隆一个特定的基因是一种通俗的说法,它的意思是形成一个含有一种特定基因的克隆。

行表达所要求的条件进行介绍。

### 用限制性内切酶切割大供体分子获得特定 DNA 片段

特定的噬菌体基因克隆到质粒中的技术对于噬菌体基因表达的调节的研究有很大的帮助。如果首先纯化出含有基因的片段,那么,克隆噬菌体基因的工作就便利得多。图 20-4 中示出一个特定噬菌体的限制性内切酶酶切图谱;假设有人想要克隆一个已经知道是处于片段 C 中的基因。一种办法是把完全消化的酶解样品加到含切割过的载体的溶液中,允许这两种 DNA 退火,然后,搜寻含有携带着所感兴趣的基因的质粒的细胞。因为 DNA 有五个片段,所以在含噬菌体 DNA 的质粒中,最多也只能有  $\frac{1}{5}$  具有所

感兴趣的特定基因。在某些情况下,值得去增加这个部分。为达到增加此部分的目的,可以先纯化片段 C,将这个片段与切割过的载体混合,再进行退火。分离一个片段最直接的办法,是在电泳分离后取出含有这个片段的条带,再对其进行抽提而得到纯化的某片段。具体做法如下:把限制性内切酶消化样品加到凝胶上,电泳一段时间,关断电源,将凝胶取出,浸泡于含有溴化乙锭 (ethidium bromide) 的缓冲液中,DNA 条带就可显现出来[也

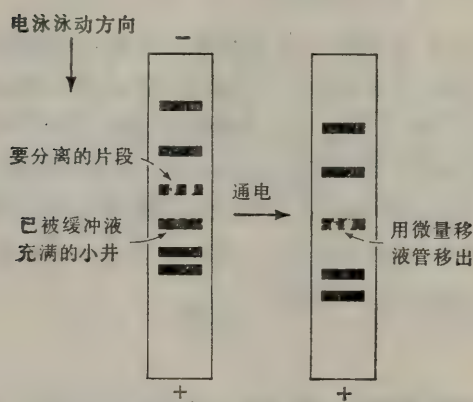


图 20-11 分离一个特定限制性内切酶片段的方法。

可见图 20-3 (b)]。因为当溴化乙锭嵌插到 DNA 分子中后,它的荧光显著地加强了。因此,使用近紫外波段的光线照射凝胶,并由发出很强荧光的位置,对 DNA 条带的存在进行定位。然后在凝胶上于所感兴趣的条带的前端挖一个小井,向小井内注满缓冲液。再通电,随之,可以见到片段在向前移动。当所感兴趣的片段移动到含缓冲液的小井时,关闭电源。这时缓冲液中含有 DNA 片段。将含 DNA 片段的缓冲液吸出,即得到纯化了的片段 (见图 20-11)。

### 重组 DNA 分子的检测

从理论上讲,如前面章节所述,DNA 分子的连接是一个直接的过程。但实际情况要复杂一些。如果将已由限制性内切酶切割过的载体,与一组未经分级分离的 DNA 的限制性内切酶片段一起退火,结果会形成好几种类型的分子,如不带有任何外来 DNA 片段的载体的自我连接的分子,带有一个或多个外来 DNA 片段的载体,以及外来 DNA 片段的自我连接的多聚体。第三类仅由外来 DNA 形成的多聚体分子不能稳定地转化已由  $\text{CaCl}_2$  处理过的细胞。因为一般地说,这类分子缺乏复制起点及复制基因。形成这类分子,除了消耗外来 DNA 片段的不利之外,并不造成对重组分子检出的任何困难。然而,在工作中确实需要用一些方法去检测经  $\text{CaCl}_2$  处理过的转化细胞中的质粒 (或噬菌体) 是否的确携带了插入的外来 DNA 片段。更进一步说,即使质粒携带了外来的供体 DNA



片段，也并不保证所带的片段恰好就一定是所感兴趣的片段。因此，进一步需要的检测手段应当是用于检测是否含有目的质粒的方法。同样地，以噬菌体及病毒为载体的工作，也存在着这种任务。本章将要介绍几种解决这种问题的方法。

保证质粒载体含有插入 DNA 片段的方法

pBR 322 是 DNA 克隆中使用得最广泛的质粒，它是一个较小的 DNA 分子，分子量为  $2.9 \times 10^6$ ，有两个不同的抗生素抗性记号，即抗四环素 ( $tet^r$ ) 及抗氨苄青霉素 ( $amp^+$ )。因此，可以通过把一个转化了 DNA 的细菌培养物接种到含有四环素或氨苄青霉素的培养基上，并置于生长温度条件下，凡在这种培养基上能生长的菌一定含有上述 pBR 322 质粒，利用此特性可以检出 pBR 322 质粒。又因 pBR 322 具有七个不同的限制性内切酶的切点，外来 DNA 可以插到这些位点上。

检测插入作用的一种最普通的方法是插入失活作用 (insertional inactivation)，这个

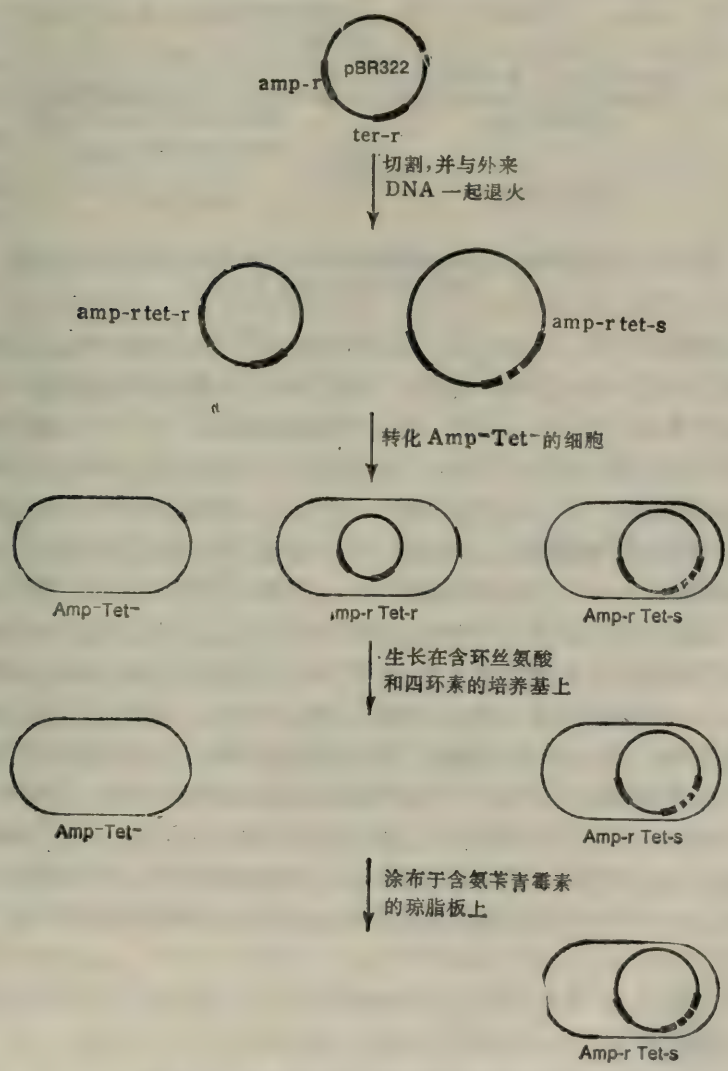


图 20-12 分离含有外来 DNA 片段的插入失活作用。

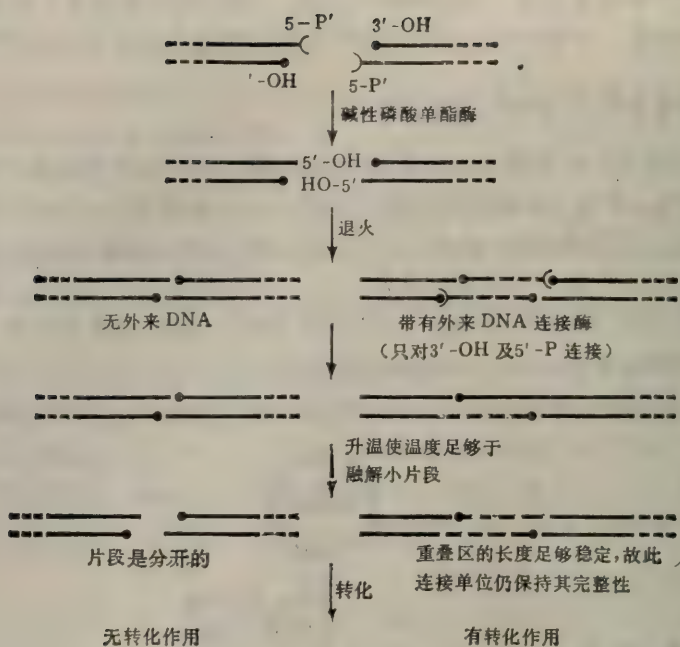


图 20-13 防止限制性内切酶切割后的质粒重新再重组到一起的方法：用碱性磷酸单酯酶处理质粒 DNA。

过程的进行示于图 20-12。BamHI 及 Sall 的位点处于 tet 基因之中。因此，在这两个位点插入外来 DNA 片段后，就产生具有  $amp^+ tet^-$  特点的质粒，因为 tet 基因（指对四环素有抗性的基因）已经失去活性。野生型细胞（ $Amp^- Tet^-$ ）不能在含有氨苄青霉素或含四环素的培养基上生长。如果一个由限制性内切酶切割而产生的 DNA 片段与一个经限制性内切酶切割过的 pBR 322 质粒一起退火，然后用这个连接的杂交质粒去转化细菌细胞，再将此种细菌细胞悬浮液涂布于含有氨苄青霉素的琼脂平板上。存活下来的细菌必然是  $Amp^+$  的，它们必定具有上述质粒。又因为 pBR 322 还携带着  $tet^+$  等位基因，因此，除非有一段外来 DNA 插到  $tet^+$  等位基因上，否则这种  $Amp^+$  的菌落也必然有  $Tet^+$  的特性。那些表现为  $Amp^+ Tet^-$  的细菌细胞则不仅有 pBR 322 质粒，而且此质粒的 tet 基因上还插有外来供体的 DNA 片段。这段操作可以通过在将细菌悬液涂布之前，让细菌在含有环丝氨酸及四环素的培养液中生长一段时间而得到简化。环丝氨酸杀死生长着的细胞。但是  $Tet^+$  细胞仅仅是被四环素抑制着，并不被四环素杀死。因此，在这个培养液中， $Tet^-$  细胞（它们生长着）被杀死了，而  $Tet^+$  细胞（它们被抑制着）没有被杀死。把接受过这种处理的细菌悬浮液再涂布于含有氨苄青霉素的培养基平板上，产生  $Amp^+ Tet^-$  的细菌菌落，所有这些菌落都具有携带着供体 DNA 片段的 pBR 322。

pBR 322 质粒中的 PstI 位点处于 amp 基因中，因此，插入失活作用也可以用在在这个位点上。如果确实进一步这么做了，就应当选择  $Amp^- Tet^-$  菌落。

EcoRI 是一种经常使用的限制性内切酶，但是 pBR 322 的 amp 基因和 tet 基因上



都没有 EcoRI 的作用位点,为此人们用重组 DNA 技术组建了一个新的质粒 pBR 325。pBR 325 中有一个抗氯霉素的抗性基因, EcoRI 酶的作用位点就在这个抗性基因内。因此,这个质粒的 EcoRI 位点的插入失活作用,可以通过细胞失去对氯霉素的抗性而测出。按上述设计已构建出许多普遍使用的有插入失活作用的质粒。对它们的介绍可见本章后面的参考文献。还有三种方法被选用来阻止无外来片段的质粒重新组装到一起。其中的一种方法就是用碱性磷酸单酯酶 (alkaline phosphatase) 作用于那些已经被限制性内切酶切割的质粒。碱性磷酸单酯酶移去了 5'-P,留下了 5'-OH 基。除非插入外来的 DNA 片段,不然,经过这种处理之后的质粒不可能再度形成有功能的质粒,如图 20-13。第二种方法,用均聚尾法进行连接自动地选择,从而防止再度组成质粒。因为被切割过的质粒的两个 3'-OH 末端都接上了由同样的碱基组成的均聚尾。方法之三是利用特殊类型的质粒,叫做柯斯质粒 (cosmids, cos-site-carrying plasmids),这个方法将在下面介绍。

柯斯质粒是一种新颖的载体,是携带着 cos 位点的质粒。它结合了质粒及  $\lambda$  噬菌体两者的特性,以便增加选择出携有外来基因的重组质粒的概率。一种典型的柯斯质粒是同时携带着利福霉素抗性基因 (rif<sup>r</sup>) 和  $\lambda$  噬菌体的 cos 位点两者的环状的 ColE1 质粒(见第十七章)。

在柯斯质粒上有两个 Hind III 的作用位点,它们把 rif 基因和 cos 位点及 ColE1 区分开来。

在第十五章已经介绍过 cos 位点切割体系,它对  $\lambda$  噬菌体 DNA 装入噬菌体头部有应答反应。如果能够满足下面两个条件:(1) DNA 分子必须含有两个 cos 位点,(2)两个 cos 位点之间的距离一定不小于 3.8 kb,不大于 4.5 kb 时,Ter 体系就可以作用于 DNA 分子。柯斯质粒可以用作载体是基于  $\lambda$  DNA 的体外包装技术,这项技术的要求就是刚才已经提过的 Ter 体系的要求。

图 20-14 中的柯斯质粒和它的 Hind III 消化产物都不能用作体外包装 (in vitro packaging)的底物,因为其 DNA 仅有一个 cos 位点。图中(a)行表示如果这种分子与另一种分子一起退火,在许多个最终分子中总会有一个二聚体(或高一些的多聚体)含有两个 cos 位点。但是 Ter 体系仍不能在这个分子上起作用,因为 cos 位点相距太近,只少数几个千碱基对。(b)行表示如果有另一种有机体的 DNA (如骆驼的 DNA)的 Hind III 限制性内切酶片段与切割过的柯斯质粒的 DNA 混合,允许它们再退火,那么将出现许多不同于(a)行中产生分子,这些分子中一定会有一种两个 cos 位点的距离适合于  $\lambda$  包装体系,其两个 cos 位点由足够量的外来 DNA 片段所分开的线性二聚体分子。因此,这种体外包装过程有可能产生出一群转导颗粒,这些颗粒都含有一条线性的、在每一端都有一个正常  $\lambda$  噬菌体 DNA 的粘性末端的、重组的柯斯质粒 DNA 的分子。注意,这种包装仅仅发生在柯斯质粒含有外来 DNA 片段的情况下(虽然也有可能装入含有多个 rif 基因的拷贝的多聚体)。(c)行表示用带有柯斯质粒的转导颗粒去感染 Rif<sup>r</sup> 型的大肠杆菌细胞的结果:不再形成噬菌体,因为转导颗粒中已没有噬菌体基因。柯斯质粒 DNA 被注射到细菌细胞中,通过  $\lambda$  DNA 的 cos 位点粘性末端而环化。柯斯 DNA 一旦被环化并被连接完全,它就可以使用 ColE1 的复制体系进行复制。将这种受感染的细菌接种到含利福霉素的琼脂培养基上,选择含有 rif 基因、ColE1 区及外来 DNA 的细胞。

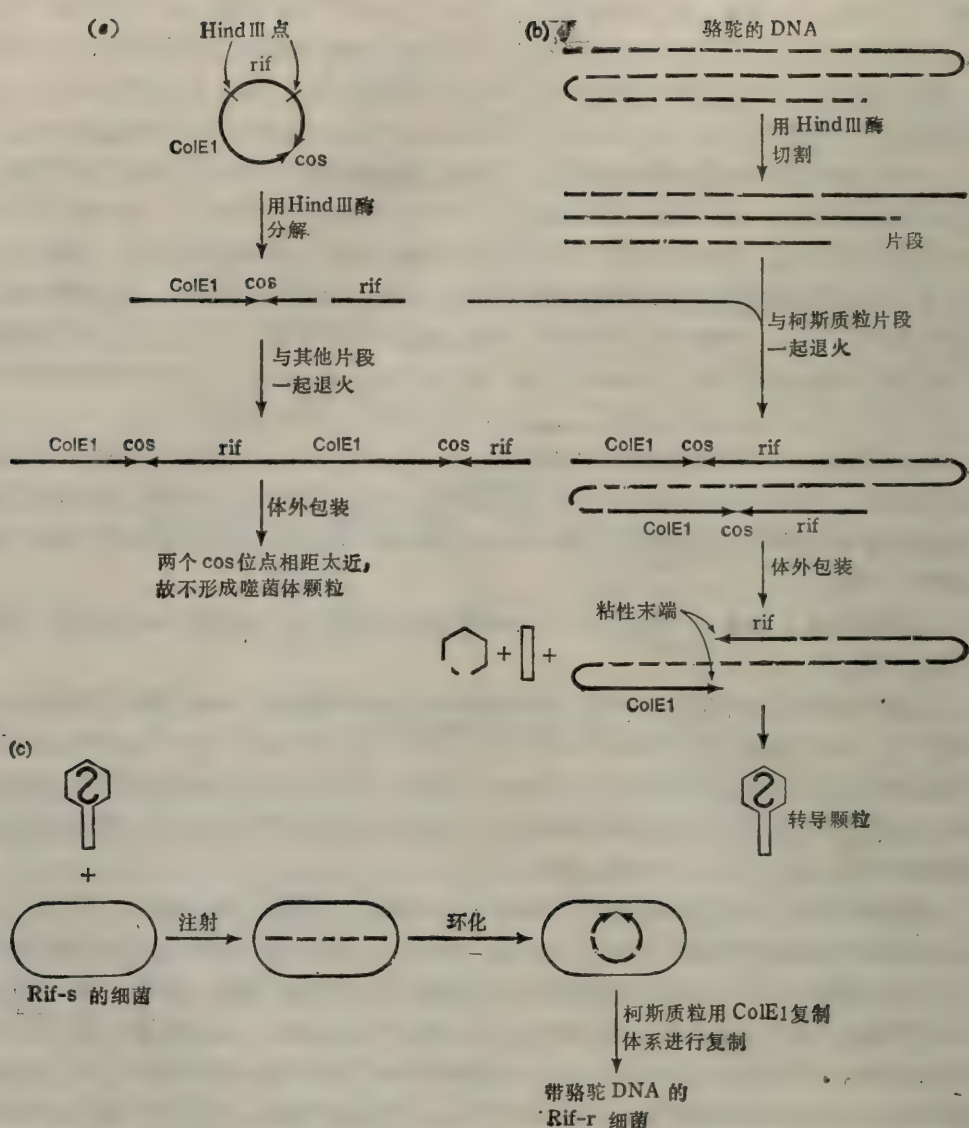


图 20-14 用柯斯质粒作为质粒 (a) 切割柯斯质粒。(b) 形成转导颗粒。(c) 转导作用。

### 保证噬菌体载体含有外来 DNA 片段的方法

有两种方法可用于克隆基因到大肠杆菌的  $\lambda$  噬菌体中。这两种方法都是基于  $\lambda$  噬菌体 DNA 有一些为溶菌生长途径所不需要的中央区基因(即 b2、int、xis、red、gam 及 cIII 基因等)。这些基因被称之为非必需区域的基因(见第十五章)。

$\lambda$  的变种  $\lambda$ gt4 $\cdot\lambda\beta$  在近非必需区域的末端有两个 EcoRI 限制性内切酶作用位点(野生型  $\lambda$  有五个 EcoRI 位点,但在制作  $\lambda$ gt4 $\cdot\lambda\beta$  时,用遗传操作法去除了三个位点)。使用这个  $\lambda$ gt4 $\cdot\lambda\beta$  载体时,先用 EcoRI 酶作用于载体 DNA,再电泳,分离出三个片段,弃去中间的片段,保留两个末端片段(图 20-15)。这两个末端片段可以通过互补的



单链末端而连接到一起，但是连接后的分子没有感染性，因为此分子的长度只有野生型  $\lambda$  DNA 的 72%，而可以装到  $\lambda$  噬菌体头部的最短 DNA 长度应当是野生型  $\lambda$  DNA 的 77%。然而，如果插入外来 DNA 片段到两个末端片段之间，产生出的杂交 DNA 分子就有感染性了，即能产生子代噬菌体，因为杂交 DNA 已可以包装到  $\lambda$  噬菌体头部。因此，如果用两个末端 DNA 片段和外来 DNA 片段的混合物进行退火而得到的 DNA 去感染大肠杆菌细菌培养物，所产生的任何噬菌体都含有外来 DNA。

$\lambda$  噬菌体的另一个变种含有 lac 操纵子（前已提到，是指乳糖操纵子）的一部分，其中包括 lacZ 基因、lacP 启动子及 lacO 操纵基因。在 lac 操纵子部分有一个限制性内切酶的切割位点，所以当外来基因在噬菌体中的这个位点上插入时，就产生出一个有组成型合成  $\beta$ -半乳糖苷酶性质的噬菌体。有几种  $\beta$ -半乳糖苷酶的底物，它们是非诱导型的底物，能在  $\beta$ -半乳糖苷酶的切割下产生有色化合物。如果把这种底物掺入到培养基琼脂中，再涂布接种 lac O<sup>+</sup> 细胞及 lac O<sup>-</sup> 细胞（构成型）的混合悬液于琼脂平板上，含 lac O<sup>-</sup> 的菌落将是有颜色的，因为只有这些细胞合成  $\beta$ -半乳糖苷酶。同样，如果把生色底物掺入到培养基琼脂中，在其上允许长出布满 lac O<sup>+</sup> 细菌的菌苔，再涂布接种 lac O<sup>+</sup> 噬菌体及 lac O<sup>-</sup> 噬菌体的混合液于 lac O<sup>+</sup> 菌苔上，lac O<sup>+</sup> 噬菌体将产生无色噬菌斑，而 lac O<sup>-</sup> 噬菌体（它含有插入的外来 DNA 片段）将产生有色的噬菌斑。

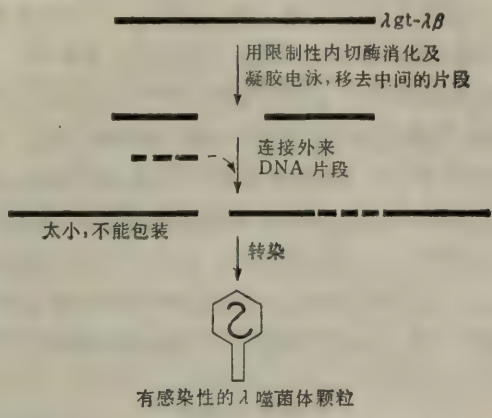


图 20-15 重组 DNA 包装到  $\lambda$  噬菌体颗粒中。

### 鉴定含有外来 DNA 片段的质粒的物理学方法

对于某些实验来说，并不需要特定的外来 DNA 片段，然而需要某些片段。例如，在某些真核 DNA 的遗传学结构的研究中需要相互独立的 DNA 区域的克隆片段，但是并不需要携带特定基因的片段。这样的质粒也许能因为它的分子量大于野生型的而易被检出。如果插入一个重组从某个纯 mRNA 制备得到的独特的 cDNA，这时分子量上特殊的增加就也是一种很有用的标准。

直接表示分子大小的方法是分离质粒 DNA，测量其长度。一般，直接用电子显微镜技术扫描是令人乏味的，因此已用简单的电泳技术取而代之，这在第十七章及图 17-4 中已经介绍过。取含有质粒的细胞的单菌落培养物的溶菌体进行电泳（通常一次同时进行 16 个不同的菌落电泳）；电泳中，每一个单菌落样品中的质粒 DNA 的量足够作为一个单个的条带而被观察到，它泳动在染色质的前方。质粒 DNA 移动的距离与它的分子量相关，质粒 DNA 的分子量愈大，它移动的距离就愈短。因此那些含有供体 DNA 的较大的质粒就移动得比那些不含供体 DNA 的质粒略慢一些，于是含有供体 DNA 片段的菌落就这样很容易地被鉴定出来了。

如果认为质粒含有外来 DNA 的一个特定的片段，又有 mRNA，就可以采用两种不

同的杂交手段去鉴别证实外来特定 DNA 片段的存在。第一种方法是,如果使用放射性 mRNA,而且此 mRNA 又只与外来的目的 DNA 片段互补,那么让含有这个质粒的细菌的总 DNA 变性后固定于硝酸纤维素滤膜上,再加入放射性 mRNA,在 DNA 与 mRNA 互补复性之后,洗涤滤膜,滤膜上存在有放射性,表明质粒中已含有互补于 mRNA 的、待测的外来 DNA 片段。如果这种 mRNA 也与宿主 DNA 的一部分互补(例如,一个大肠杆菌的基因克隆到大肠杆菌中),那么在进行杂交实验之前,必须先把质粒 DNA 提纯出来,使质粒 DNA 与放射性 mRNA 杂交,以找出质粒中与 mRNA 互补的那一段外来 DNA 片段。第二种方法是用 R 成泡法(R-looping)。如果已得到纯质粒 DNA(见图 20-16),可以使用一种精巧的电子显微镜技术:将质粒 DNA 在一含有 mRNA 的缓冲液中进行部分变性,在这种环境中,DNA-RNA 杂交体比双链 DNA 更加稳定。因此,如果存在着可以互补的 RNA,那么这种 RNA 就会侵入双螺旋,产生一个泡状物,此泡由一单链分支(DNA)及一双链分支(DNA-RNA)所组成。

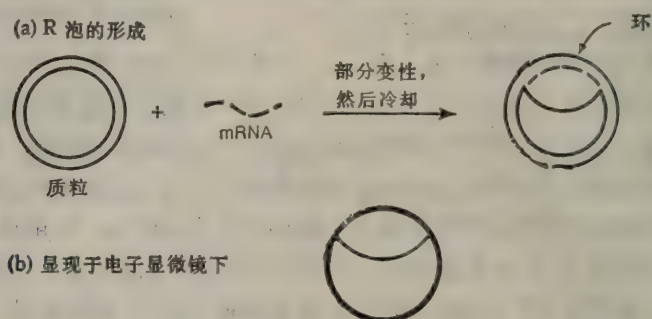


图 20-16 R 泡的形成。因为环状 DNA 分子总是超螺旋的,在变性之前,单链切口(未指出)通常被引进质粒中,因此质粒分子将较为舒展并更容易被观察到。

### 对于含有特定 DNA 片段的质粒的集群筛选

检测特定外来基因最简便的方法是互补作用。最早报道的实验是检测酵母组氨酸合成中的一个特定基因,是通过质粒与酵母染色质片段重组,再把一群重组 DNA 分子转化到一种特殊的  $\text{His}^-$  大肠杆菌突变株中而检测得到的,转化后有些细胞就含有了酵母基因,然后在缺乏组氨酸的培养基上挑选能生长的  $\text{His}^+$  菌落。但以上方法只对那些可以互补的基因有用,对于被克隆到细菌中的动物基因因此法一般不成功。因为通常动物 DNA 在细菌中或是不表达,或是没有相应的细菌基因。这一部分将要介绍两种较常使用的方法。

一种是菌落或原位杂交检验法(colony or in situ hybridization assay),此法可用于检验是否存在特定基因(图 20-17)。另一种是免疫化学检验法,此法可检验是否存在特定基因产生的人们所感兴趣的蛋白质产物。

在第一种原位杂交法中,先把被测定的菌落从培养皿中琼脂培养基表面转移到滤膜上。每一个菌落的一部分遗留在琼脂上,这一部分构成参考皿。滤膜用 NaOH 处理,NaOH 溶解细菌,释放变性 DNA,然后,烤干这张滤膜,变性的 DNA 就被固定于滤膜上。将此烤干了的滤膜再浸于  $^{32}\text{P}$  标记的、与被搜索的基因互补的 mRNA 的溶液



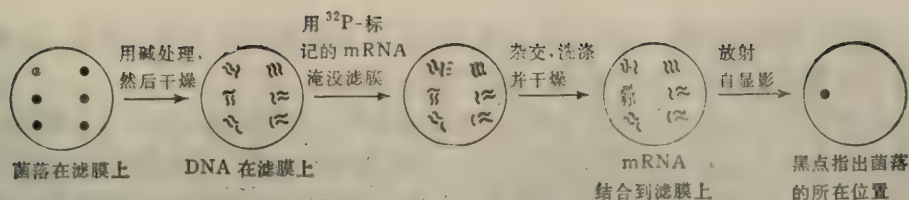


图 20-17 菌落杂交。图中未表示出参考皿, 滤膜上的菌落就是由参考皿复印而来。

中, 并置于最良好的复性条件之下, 使 mRNA 与 DNA 杂交。然后洗涤滤膜以除去未结合的  $^{32}\text{P}$ -mRNA; 遗留在滤膜上的结合的放射性仅仅是那些被杂交了的  $^{32}\text{P}$  mRNA 的放射性。将滤膜晾干, 并放到自显影底片上, 底片经显影后出现的黑点就是我们感兴趣的菌落的位置, 然后可以从参考皿上挑出这个菌落。

对于噬菌体载体也可以用类似的方法。噬菌体被接种到细菌菌苔上, 当形成噬菌斑时, 将滤膜置于琼脂上, 来自每个噬菌斑的噬菌体粘附到滤膜上, 然后用碱如上法处理滤膜, 以使 DNA 固定并变性。用上述杂交过程来检定含有人所感兴趣的基因的噬菌体所形成的噬菌斑, 然后从参考皿上将目的噬菌斑移出。

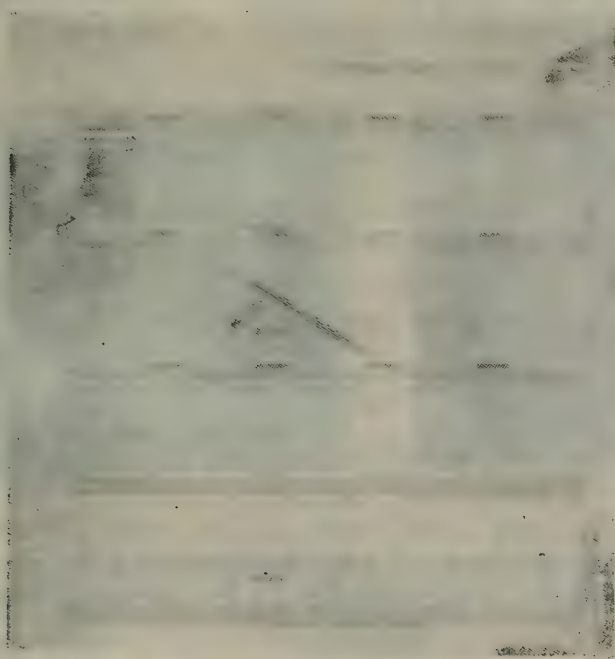


图 20-18 放射性抗体试验。(a,b) 涂有抗胰岛素抗体的塑料圆形薄膜接触到培养皿中菌落的表面上。(c) 从产生胰岛素的菌落中形成的胰岛素分子结合到抗体上。(d) 塑料圆膜浸入放射性抗胰岛素的抗体溶液中。(e) 放射性粘着到塑料圆膜上产生胰岛素的菌落处。

另一种是免疫检验法。如果形成了目的基因的蛋白质产物, 就可以采用两种免疫学技术来鉴定产生蛋白质的菌落。如果蛋白质被分泌出来, 就使用放射性抗体试验 (图 20-18)。在这个试验中, 将一张涂有能与基因的蛋白产物相结合的抗体的圆盘状塑料薄膜压到菌落上面。如果菌落上存在着蛋白质产物, 这种蛋白质就会因抗原-抗体关系结合到薄膜上, 然后把这张圆盘状塑料薄膜再浸入到一个含放射性抗体的溶液中, 放射性抗体

就结合到塑料圆盘状薄膜上的蛋白质上,此圆盘经过洗涤后,进行放射自显影。放射性的位置表示了能合成基因产物的菌落所在的位置。

在某些情况下,插入的基因可取代一个已在质粒中编码的蛋白质的终止密码子(即质粒基因本身的终止子),并把这个插入的、人们感兴趣的基因偶联到质粒基因上去。这样的结果是产生出一个称之为融合蛋白质(fused protein)的杂交蛋白质分子。这个杂交蛋白质分子带着一部分质粒蛋白质的序列,起始于氨基末端,终止于所感兴趣的蛋白质序列的羧基末端。前面介绍的免疫学方法通常也适用于这种融合蛋白质。

在某些情况下,融合作用有特殊的意义。如正常情况下细菌不把人们感兴趣的某蛋白质分泌出来,就可将目的基因插入到同时也编码 $\beta$ -内酰胺酶的pBR 322载体上。一旦所感兴趣的基因被插入载体,它的基因产物某蛋白质就与 $\beta$ -内酰胺酶融合在一起。 $\beta$ -内酰胺酶是可以分泌的,它携带着人们感兴趣的某蛋白质穿过细胞壁,来到细胞外,这时,就可以用上面介绍过的放射性抗体法或下面将要介绍的免疫沉淀法进行化验。在检验蛋白质产物是否出现之前,可以根据pBR 322载体对四环素的抗性来选择Tet<sup>-</sup>细胞(这些细胞都含有pBR 322),并通过插入的失活作用来检定是否有外来DNA片段插到载体上。即通过在Tet<sup>-</sup>细胞中选择Amp<sup>r</sup>特性,可以把Tet<sup>+</sup> Amp<sup>r</sup>细胞选出,再用上述方法或下面将介绍的方法对这些被选中的细胞作进一步的蛋白质产物的检验工作。

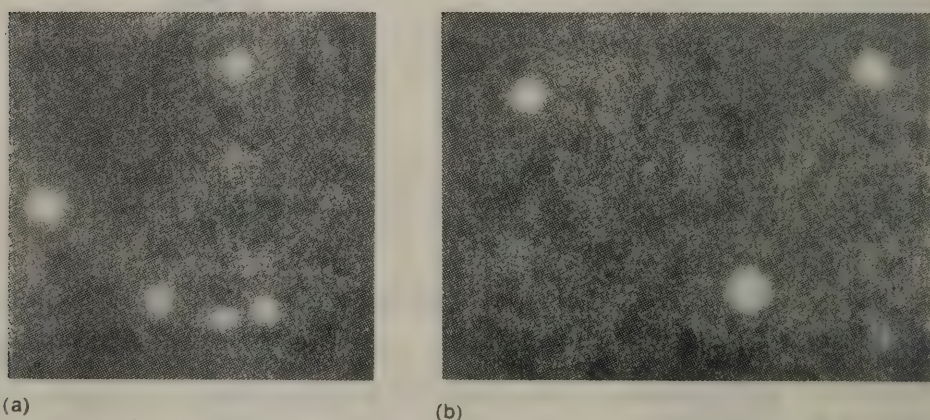


图 20-19 免疫沉淀试验。在含有抗- $\beta$  半乳糖苷酶抗体的琼脂培养基上形成产生 $\beta$ -半乳糖苷酶的菌落。(a) 围绕着每一个菌落形成沉淀素区带。(b) Lac<sup>+</sup> 菌落(有沉淀素区带)及 Lac<sup>-</sup> 菌落(无区带)的混合物。

还有一种是免疫沉淀检验法(immunoprecipitation)。此法也可以用来鉴定菌落是否产生蛋白质。如果菌落分泌蛋白质,那么被称之为沉淀素(precipitation)的抗体-抗原复合物将围绕在菌落周围,在含有抗体的培养基上形成一个用肉眼可以看得见的白色的沉淀环(见图 20-19)。将上面这个免疫沉淀法作一些微小的修改,就适用于测定那些并不分泌到细胞外的蛋白质。第一种修改是选用对于噬菌体 $\lambda$  cl 857 的溶源菌为宿主细胞。 $\lambda$  cl857 是含有温度敏感阻遏物(heat-sensitive repressor)的 $\lambda$  突变株,它的溶源性在温度升高到 42°C 时失去。测试过程为:先把含有待测菌落的主培养皿中的菌落经复印法接种到一个含有抗体的琼脂的复印皿中,37°C 培养,当复印皿中开始可以看到菌



落时,将复印皿移到 42℃ 下培养。大约 1 小时之后,温度升到 42℃,诱导许多细胞进入溶菌过程,细菌细胞溶解,因而释放出细胞内含物。于是所感兴趣的蛋白质得以与琼脂培养基中的抗体反应,在菌落周围生成沉淀环。第二种修改是把含有抗体和溶菌酶的温热的琼脂培养基倒在菌落上,待培养基冷却变硬,溶菌酶在细菌表面逐步地溶解细胞,细胞内的蛋白质逐步地释放出来。同样地,在人们感兴趣的菌落周围形成了沉淀环。

## 基 因 文 库

许多实验室反复利用重组 DNA 技术从一种单一的生物中分离特定的基因或 DNA 片段。例如,已经通过对一族克隆了的 DNA 片段进行分析,研究了果蝇(*Drosophila*)中基因的空间结构 (spatial organization),这部分将在第二十二章讨论。要完成一个所需要的 DNA 片段的完全的克隆过程是很花费时间的。为此,已经建立了含有杂交质粒的细菌集群,这个细菌集群有足够的容量,因此细胞总 DNA 的每一个部分都在集群中可出现一次 (有时可出现多次)。这个集群被称为菌落银行(colony banks) 或基因文库 (gene library)。在这一部分我们将讨论如何建立基因文库,以及如何测定能反映每个序列所需要的克隆数。

酵母 DNA 文库已在加利福尼亚大学的 Santa Barbara 分校建立。载体是 ColE1 质粒。为了形成文库,可用 EcoRI 酶切割 ColE1 DNA, EcoRI 酶在 ColE1 质粒上有一个切点。做这个工作要避免从文库中得到的许多基因是以一个完整的单位存在,但是用限制性内切酶又会把某些基因切开。因此不要用限制性内切酶作用于酵母 DNA。而是应当采用流体剪切力无序地打断(通过迅速地搅拌酵母 DNA 溶液)。为了把酵母 DNA 片段连接到 ColE1 质粒,采用了多聚尾连接法,即把多聚 (dT) 尾加到切开的 ColE1 DNA 的 3'-OH 末端上,并在酵母 DNA 片段上加多聚 (dA) 尾,然后进行分子连接,再转化细菌。含有 ColE1 后的细菌细胞对加到培养基琼脂中的细菌素 E1 有抗性,因此可以利用这个特性很容易地把含有 ColE1 质粒的细胞分离出来。再把分离到的菌落——深刺接种到一个塑料板上的各个小井中。在每一个这种塑料板上有 96 个小井,每个小井中都充满着固体的储存培养基。将塑料板储于 -80℃。在 -80℃ 的温度下,于特殊的储存培养基中,细菌大约可以维持存活的状态。这个塑料板中包含了一个完整的基因文库。现已观察到至少有 70% 的菌落中含有酵母 DNA。

下面给出简单的计算式,使人们可以测出为形成一个完全的文库需要多少个菌落。

假定现在有一个 DNA 分子,这个分子含有一些一定大小的 DNA 片段。现在用  $f$  表示每一个片段占该有机体总 DNA 的组分数,用  $N$  表示集群中的菌落数目,用  $P$  表示某特殊基因在这个集群菌落中出现的概率,于是

$$P = (1 - f)^N$$

或

$$N = \ln(1 - P) / \ln(1 - f)$$

假设我们希望概率达到 0.99, 即每一段序列都在文库之中,即  $P = 0.99$ 。如果供体 DNA 是单倍体酵母 (其分子量约为  $10^{10}$ ), 而且剪切出的 DNA 片段的平均分子量为

$8 \times 10^6$ , 那么

$$f = 8 \times 10^6 / 10^{10} = 0.0008$$

于是

$$N = [\ln(1 - 0.99)] / [\ln(1 - 0.0008)] = 5754$$

因此,当文库中大约有 5 800 个菌落时,就有 99% 的把握说,任何一个酵母基因至少座落于一个菌落之中。如果要进一步细究,可随意挑选少数菌落,酵母 DNA 的序列在它们中很可能是重叠的。

对于具有大 DNA 分子的细胞(如果蝇)来说,产生基因文库所必需的菌落数约为 30 000 个。如加大构成 DNA 片段的大小  $n$  倍,那么菌落数约可减少  $n$  倍。

如果需要一个含有特定 DNA 片段的克隆,在原则上可以采用本章前面部分所介绍的任何一种方法进行筛选。

也可以用  $\lambda$  噬菌体为克隆载体建立基因文库,通过对保存在多井塑料板中的重组噬菌体进行悬浮可以得到文库中的多个基因。噬菌体文库插进到单个噬菌体颗粒中的 DNA 的最大容量是有限的。对于一种有机体来说,建立一个完全的文库所需要的最少的噬菌体数目,可以由允许插到单个噬菌体中的外来 DNA 的最大容量而测定出来。对已有的  $\lambda$  载体来说,插入的 DNA 片段分子量不能超过  $8 \times 10^5$ 。因此建立一个噬菌体文库所需的最少噬菌体数目对于大肠杆菌、酵母和果蝇来说,分别是 2 900、5 800 及 300 000。

## 从克隆的基因生产蛋白质

基因工程的目标之一是大量地生产一些难于取得的特殊蛋白质(例如,在每个原有细胞中仅有很少数几个分子的某蛋白质;或是某种本身有致病性的蛋白质,其操作一般是有毒的)。遗传工程方法从原理上讲是很简单的:把基因偶联到有启动子的载体上,进行测试以保证基因在正确的方向上,因此编码链被转录,进而被翻译成蛋白质。所选用的载体通常是一种高拷贝数的质粒,或是一种活跃复制的噬菌体(如  $\lambda$ ),因此细胞中存在着大量拷贝的基因。这就能使细胞合成出高达细胞总蛋白质含量的 1—5% 之多的基因产物——某特定蛋白质,或以每升培养基形成 10—50 mg 蛋白质的产量生产这种特定的蛋白质。

实际上,不会总是遇到生产大量目的蛋白质的目标,因为还有各种各样的理论的和技术性的问题。下一部分将要介绍某些这类问题以及如何解决这些问题的种种办法。

### 生产真核生物蛋白质中的问题

如果要得到细菌基因或噬菌体基因产生的蛋白质,可以用直接的方法生产它们。但如果是真核基因,就会遇到下列一些特殊的问题:

1. 可能细菌的 RNA 聚合酶不能识别真核的启动子。
2. 由真核基因转录生成的 mRNA 可能缺乏结合到细菌核糖体上的 Shine-Dalgarno 序列。
3. 真核 mRNA 可能含有必须除去的内含子。
4. 真核基因的产物通常必须再进行加工处理。



5. 细菌本身具有的蛋白酶常常识别出由细菌合成的真核蛋白质是外来蛋白酶,而加以破坏。

其中第1、第2个问题通常可以通过把基因偶联到 lac 或  $\lambda$  的启动子上而解决,下面将要叙述。内含子的问题是可以避免的。如若我们可以从细胞中获得充分加工过的 mRNA,那么可以由它来制备 cDNA,再把这种 cDNA 连到载体上就可以回避内含子问题;另一种做法,如若已知蛋白质分子并不很大,其氨基酸序列又已弄清楚,那么,可以直接合成该蛋白质的基因,再把此合成基因连接到载体上。关于加工问题相对地说比较难解决。虽然在某些情况下,先分离出新生蛋白质,然后可以在体外对它进行加工,然而许多成熟的蛋白质是初始翻译产物的修饰形式,这个事实对于使用细菌来生产真核蛋白质有一定的限制作用。细菌蛋白酶对真核蛋白质的破坏则推测可以通过突变作用而加以消除,至少部分地消除。下面还将简短地介绍一些用细菌合成的动物蛋白质。

近来,酵母的 2 微米质粒(见第十七章)已经用作真核基因克隆的运载物了。到目前为止已经进行过的研究表明,对于酵母这个体系来说,并不存在前面所述的真核基因在细菌中表达可能遇到的种种困难及问题,或者说这些问题在酵母体系中已得到解决。酵母的 RNA 聚合酶能识别动物基因的启动子,酵母中的加工体系显然可以移去动物基因中的内含子并对某些动物蛋白质进行加工。还有,在酵母体系中,外来的真核蛋白质降解得较少,这一特点很令人注目,因为也许许多真核蛋白质和原核蛋白质有着至今还未被人们认识的某些很不寻常的性质。酵母体系在 1982 年之前,还没有被适当地开发使用,至今已发现有众多的潜在用途。此外,上面的结果,也预示着原核生物与真核生物之间的进化过渡(evolutionary transition)是一个有重大意义的生物化学事件——没有真正的细胞核的细胞与已经有真正的细胞核的细胞确实有某些同一性。

## 一个克隆基因的表达

克隆基因的表达要求基因能够转录并能够翻译。让我们来看一看人们是如何设计一个使克隆基因可以表达的体系的。

通常,限制性内切酶的位点把基因的编码区与其启动子分割开,因此,除非再提供另一个启动子,当该目的基因被插入到一载体之后,方能被转录出相应的 mRNA。如果这个载体在插入位点附近含有一个载体的启动子,那就可以利用这个启动子(通常,这个启动子应当是与插入位点相邻的那个基因的 RNA 聚合酶的结合位点)。但是,对于许多载体来讲,常常不存在这样一种启动子,于是就必须再外加一个启动子到克隆载体上。因此,进行这类工作时显然应当插入一个很强的启动子以达到最大的转录效果。最普遍的办法是用重组技术再设计一个质粒,它或是具有大肠杆菌乳糖操纵子的启动基因-操纵基因区域(第十四章)或是具有大肠杆菌  $\lambda$  噬菌体的 cI 阻遏基因-oL-pL 区域(见第十五章)。这两种体系都可以对克隆系的基因的转录作用进行调控。例如,在第一种情况下,可以通过加入乳糖操纵子的诱导物(如乳糖或 IPTG)而使转录开始,也可以通过加入葡萄糖(降低 cAMP 的浓度)而终止转录。对于  $\lambda$  体系,如果使用有温度敏感阻遏物的噬菌株(temperature-sensitive repressor, 被称之为 cI 857),当温度升高到 40°C 以上时,转录开始,在这种较高的温度下阻遏物变性;然而当温度下降到 34°C 时,转录就会终止(在该温度下,阻遏物又有活性了)。

在限制性内切酶作用的情况下,可能会遇到一种常见的情况,即酶将目的基因的编码序列与核糖体结合位点切开。因此,进行表达时还必须提供这个位点。在这方面,以上介绍的插入 lac 及  $\lambda$  体系是经过很好的设计而产生的,它们具有很强的核糖体结合位点。但是,如果将一个具有目的基因的限制性内切酶片段偶联到有功能的启动子及核糖体结合位点上并不能保证基因一定能正确地表达(图 20-20)。基因的限制性内切酶片段的两端

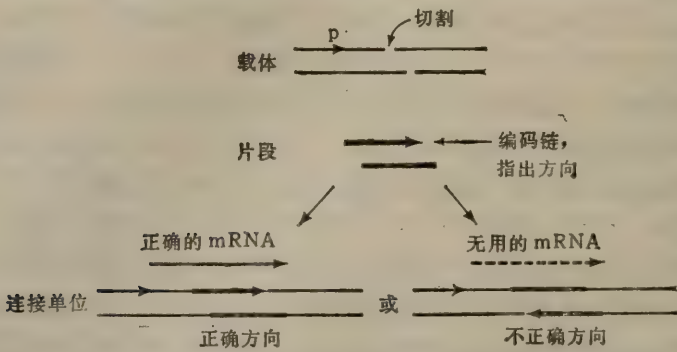


图 20-20 一个克隆基因以相对于启动子 P 的两种方向插入到载体中。

具有相同的单链末端,因此基因可以以两种可能的方向插入到载体中。鉴于任何基因只有一条链是编码链,因此,只有按某个方向插入,才能产生有用的 mRNA。两种方向的插入作用以同样的概率发生着,因此,人们必须从中选择出一种重组载体,在这种载体中基因处于正确的方向。如果有目的基因产物的合成,那么,可以很容易地检测出具有正确插入方向的克隆。然而,如果没有发生产物合成作用,那么就要应用前面部分所叙述的一些方

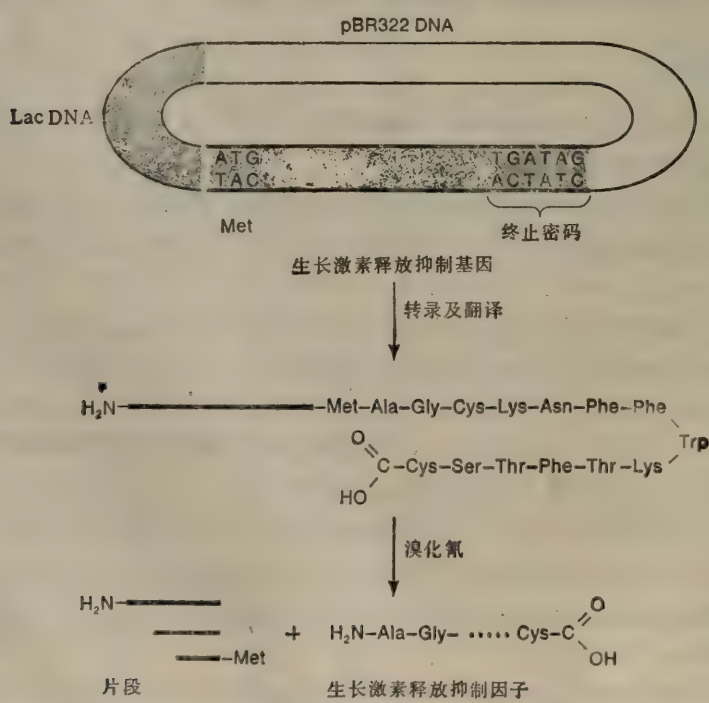


图 20-21 化学合成的生长激素释放抑制因子基因连接到 pBR322 lac 上。



法进行检测。

下面将要介绍合成真核蛋白质的两种成功的尝试。从细菌的基因表达调节以及蛋白质合成的基础研究得到许多详细的知识,我们可以看到研究者是如何运用这些知识作出贡献,取得成功的。

### 生长激素释放抑制因子的合成

生长激素释放抑制因子是在下丘脑合成的由 14 个氨基酸残基组成的一种多肽激素。这个很短的多肽已经在大肠杆菌中用一个很简便的方法合成。先通过化学合成技术合成出一个含有 51 个碱基对的双链 DNA 分子,其编码链的碱基对序列是

TAC-(42 个编码生长激素释放抑制因子的碱基)-ACTATC;

相应的 mRNA 的碱基序列是:

AUG-(42个编码生长激素释放抑制因子的碱基)-UG AUAG

因此,这个 mRNA 有一个蛋氨酸密码子(这个密码子不是用于起始作用),一个生长激素释放抑制因子的编码序列,以及两个终止密码子。所采用的载体是经过修改的 pBR322,这个修改质粒含有 lac 的启动-操纵基因和 lacZ 基因的一部分,这一部分编码  $\beta$ -半乳糖苷酶的  $\text{NH}_2$ -端片段。载体被限制性内切酶作用,在 lacZ 基因(图 20-21)的一个位点处被切开,形成平整末端;然后,化学合成的 DNA 分子用平整末端连接法连到被切割的质粒上。当乳糖操纵子被诱导时,就形成一种蛋白质,它是通过蛋氨酸将  $\beta$ -半乳糖苷酶的  $\text{NH}_2$  端的部分序列和生长激素释放抑制因子偶联到一起的一种融合蛋白质,并且被化学合成的 DNA 的两个重复终止密码子所中止。提纯这个蛋白质,再用溴化氰处理,溴化氰仅在蛋氨酸的羧基侧进行切割。后来,连接物蛋氨酸继续附着在  $\beta$ -半乳糖苷酶片段上,而生长激素释放抑制因子则被释放出来。因此,采用溴化氰对偶联着的蛋氨酸的切割技术,对从细菌蛋白质中分离出的任何一种本身并不含有蛋氨酸的融合多肽,是一种很有用的处理方法。

### 胰岛素原和胰岛素的合成

胰岛素是一种激素,它调节糖代谢,是临床上处理糖尿病患者的主要药物。正常情况下人们可从牛胰腺中获得制剂。胰岛素本身不是基因的直接产物,而是基因合成产物再通过几步后处理而形成的物质(图 20-22)。

基因的直接产物是前胰岛素原。经第一次切割后产生胰岛素原,即首先得到在两个外端片段间(在 A 及 B 片段间)具有两个二硫键的胰岛素原,然后再移去中间片段(C 片段),最终形成胰岛素。这个过程的终结是两条分别含有 21 个氨基酸残基及 30 个氨基酸残基的肽链,通过两个二硫键组成一个胰岛素分子。

胰岛素已经以两种不同的方式在大肠杆菌中合成。第一种方法类似于生长激素抑制因子合成的办法,但有些外加的步骤。首先分别合成编码 A 链的 DNA 及编码 B 链的 DNA,接着分别把这两个 DNA 片段各自融合到  $\beta$ -半乳糖苷酶的基因上,并再克隆到两个不同的细菌中。然后,分离出融合蛋白质  $\beta$ -gal-A 及融合蛋白质  $\beta$ -gal-B,再从这两个融合蛋白质中提纯出 A 链及 B 链。最后将 A 链 B 链混合,二硫键自发形成。在 A 链和 B 链的 6 个半胱氨酸之间无序地形成着二硫键,因此有活性的胰岛素的产量很低。第二种方

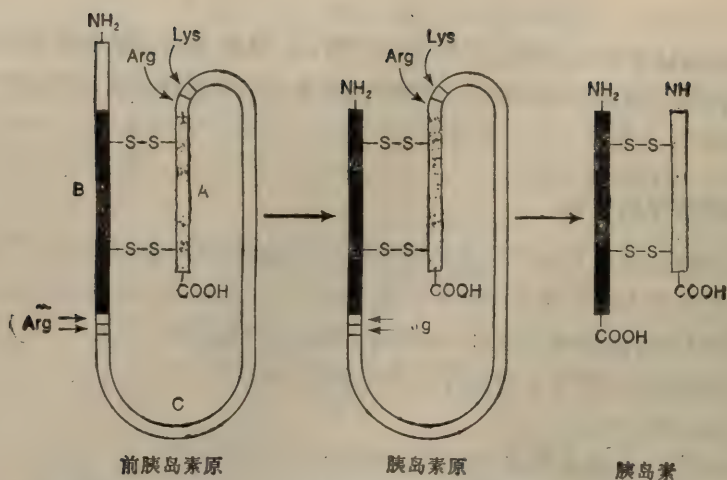


图 20-22 前胰岛素原体内转变成胰岛素原及胰岛素。书中内容讨论涉及图中四个氨基酸的位置。

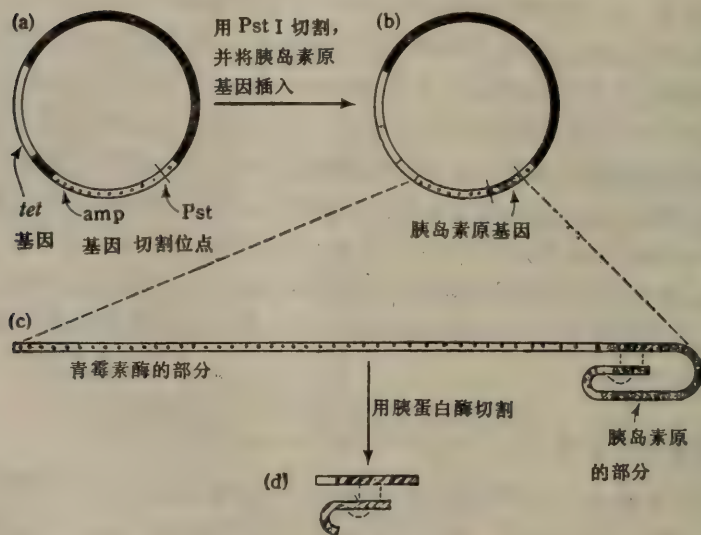


图 20-23 猪胰岛素在大肠杆菌中的生产。(a) 载体 pBR 322 质粒。(b) 含有胰岛素原基因序列的 pBR 322。(c) 杂交基因产物, 指出了胰岛素原的部分(斜线)及二硫键(虚线)。(d) 胰蛋白酶消化杂交基因产物的结果。

法,将编码胰岛素原的 DNA 克隆到 pBR 322 质粒中。这个方法首先要从猪的胰脏细胞中分离 mRNA,虽然这个 mRNA 制剂混杂着各种细胞质 mRNA,但是它富含前胰岛素原 mRNA,因为前胰岛素原是胰脏细胞的主要产物。利用这种 mRNA 加入逆转录酶合成 cDNA,接着再用限制性内切酶 Pst I 作用后得出一定的 DNA 片段,再将此 DNA 片段克隆到 pBR322 的  $\beta$  内酰胺酶中的 Pst I 位点中(图 20-23)。pBR322 的基因型是  $tet^+ amp^+$ 。上述杂交 pBR322 则是  $tet^+ amp^-$ ,对  $Tet^- Amp^-$  型的大肠杆菌培养物进行转化之后,筛选出  $Tet^+ Amp^-$ ,以保证某些 DNA 插入到质粒基因中。然后再用放射性抗胰岛素的抗体对  $Tet^+ Amp^-$  菌落进行筛选,以探测类胰岛素物质存在。找到有能力



产生胰岛素的菌落,再测定这种质粒中的 PstI 片段的碱基序列(见图 20-23),得知虽然其中存在着胰岛素原序列,但是 PstI 作用后这种片段中失去了前胰岛素原的基因的一部分。体内转录及翻译的最终产物是  $\text{H}_2\text{N}-\beta$ -内酰胺酶-胰岛素原- $\text{COOH}$  的杂交分子。这种胰岛素原分子中的 C 片段使胰岛素原能正确折叠。另一方面, $\beta$ -内酰胺酶片段是那么大,使得细菌蛋白质酶未能识别出胰岛素原是外来蛋白质,因此胰岛素原没有被蛋白酶降解而能保留下来,并在以后能发挥生物活性。

进一步胰岛素分子可以于体外从上述杂交蛋白质上被切下来。注意,在图 20-22 中近 C 片段的末端处有三个精氨酸,这是很有用的,因为胰蛋白酶仅仅在精氨酸及赖氨酸的羧基侧切断肽链,这一点可以在工程性研究中加以利用。十分短暂地处理  $\beta$ -内酰胺酶-胰岛素原的杂交分子(短促地处理使酶不至于在每一个胰蛋白酶的敏感位点上都作用,而把底物完全切成片段),在 A、C 片段之间适当的位置上临时形成一个切点,而在 B 链的羧基端有另一个切点,使之只留下一二个多余的精氨酸。另外也还需要有一个切点以便将 B 链与  $\beta$ -内酰胺酶分开。但是,在切开两者的位置上既没有赖氨酸也没有精氨酸,因此从  $\beta$ -内酰胺酶而来的少数几个氨基酸还留在 B 链,这样的胰岛素并不完美;不过,这个修饰过的分子也有生物学功能。目前还不清楚这种修饰方法所形成的产物在临床上是否有用。

### 从克隆的基因合成其他的真核生物蛋白质

在细菌中已经合成了几种其他真核生物蛋白质。酵母的组氨酸合成体系中的一个酶的基因插到  $\lambda$  噬菌体上,于大肠杆菌中有效地表达出活性。还在大肠杆菌中生成了鸡卵清蛋白。干扰素也已经可以在大肠杆菌中合成。人们对人的干扰素非常感兴趣,因为干扰素有抗病毒作用,而且可能是抗肿瘤的物质。现在每个细胞中干扰素产生的量还较低,但是是在一个经充分诱导的  $\text{lac}^+$  细胞中却可产生 100 000 拷贝的  $\beta$ -半乳糖苷酶。兔的  $\beta$  珠蛋白基因 C 也被克隆插入到 SV 40 病毒中。当用这个杂交病毒去感染猴肾细胞时,合成了兔  $\beta$  珠蛋白。

原则上,任何真核生物蛋白质都能在适当组建的细菌中合成出来,大量的研究正在逐步地取得成就。用这些方法使与医学有关的重要的物质能够人工合成出来,这种巨大的潜力说明人们有理由为之付出巨大的努力。

## 重组 DNA 技术的应用

至此,我们已经讨论了利用重组 DNA 技术生产有用的蛋白质的主要方面,这当然是极其重要的,同时,这种技术还在其他两方面很有用处:(1) 作为改变有机体基因型的方法;(2) 作为一种研究工具。

细菌可以完成大量的化学反应,但是没有任何一种细菌可以进行所有的各种反应。然而,通过遗传工程,可以把细菌有机体改造成具有好几种细菌特点的集合体。例如,由不同细菌中分离得到的好几种基因已经被插到同一个质粒中,这种质粒又被置于海洋细菌中,这样产生出来的细菌具有代谢石油的能力。现在已将这种细菌用在清除海洋中石油溢流污染的工作中。进一步,还有许多生物技术公司正在制作人工设计的、可以合成工业上许多重要的化学产品的细菌。还有,现已选用了一些有机体,它们能够生产大量的极

其昂贵的抗生素药物。其生产过程成本低得多,因此药物的售价可能会相当低廉。还有人工设计的细菌可以比较有效地分解废物,以及固氮(去改良土壤的肥力)。近来还有人以很大的努力去创造能够将生物废物转变为乙醇的细菌。这些遗传工程在世界经济、环境保护及生活水平上将会引起巨大的冲击。

改变植物的基因型也是重组 DNA 技术的重要应用。在第十七章我们已讨论过致癌农杆菌 (*Agrobacterium tumefaciens*) 及其 Ti 质粒。致癌农杆菌在质粒中产生冠瘿肿瘤。这些肿瘤致使整合了外来目的基因的质粒 DNA 进到植物染色体中,通过遗传工程就有可能把特定基因由一种植物中分离出来,引入到这种 Ti 质粒中,然后用具有杂交 Ti 质粒的细菌感染第二种植物,把基因从第一种植物转移到第二种植物中,人们企图以这种方式进行植物育种。有一个例子就是企图通过引入豆科(如豌豆,大豆)的基因改变谷物如小麦的根部表面的结构,以便使谷科植物兼有豆科植物的特点,如建立固氮菌的根瘤;如果能成功,可使生长谷物的土壤对氮肥的需要减少。

重组 DNA 技术对基础研究也有巨大的冲击。本书已经介绍了一些细菌突变株,这些突变株有简化了的体系,使得突变株细菌更加便于人们对它们进行研究(如 lac 操纵子的大量突变体)。对于一些简单的突变体可以采用标准的遗传学技术直接获得(如突变作用及遗传重组);但是对于一些需要许多种遗传记号的复杂的突变体(这些记号也许十分紧密地靠在一起)来说,因为产生突变体的频率可能很低,因而分离获得这些突变体的工作是极为庞杂而乏味的。重组 DNA 技术则可以简化获得突变体的步骤,因为可以提纯含有遗传学记号的片段并且可以在试管内使它们结合到一起。这样做可大大节省时间,并且可以组建出一些用其他方式往往不可能制成的突变体。

此项新技术对基础研究冲击最大的是真核生物,特别是真核基因的调节。在第十四章介绍的观察细菌操纵子的实验,这样的研究手段在真核生物中是行不通的。因为真核生物是双倍体,分离突变体的工作极端困难。而且,除了像酵母那样的高等单细胞生物之外,没有简单的办法可以对真核细胞生物进行多种多样的遗传学操作。

关于细菌中基因的活性可以通过在各种条件下研究 mRNA 合成而测得。最方便的测定法是 DNA-RNA 杂交反应,这种检验需要有来源丰富的含待研究基因的 DNA。用标准的遗传学方法或是用重组 DNA 技术得到的转录颗粒就是进行这类实验的 DNA 来源。例如,测定 lac mRNA 可通过它与  $\lambda$  lac DNA 的杂交。真核体系并不存在转录颗粒,但是,通过 DNA 技术可以获得,即把待研究其调节作用的真核基因先克隆到噬菌体或质粒载体中,然后让这种噬菌体生长,或让被上述质粒转化了的细菌生长,以积累 DNA,从而为研究者提供出大量的真核基因 DNA。含有这种特定 DNA 序列的载体被用作 mRNA 的探针:将这种 DNA 进行变性后,再加到细胞中,细胞内含有已被放射性标记的各种 mRNA。变性的 DNA 就可以从大量的 mRNA 中搜查到的 mRNA。上述检定 mRNA 的方法是基于其本身具有可以与 DNA 杂交的特殊能力。应用这项技术进行检测,先是把变性的 DNA 探针附着到硝酸纤维素滤膜上,然后将特定的 mRNA 再结合到滤膜上,即除了能杂交到变性 DNA 上的 mRNA 外,一般的 mRNA 并不能结合到滤膜上,这样就能检测到特定 RNA 的存在。这项技术已将真核基因调节的研究工作大大推进了一步,也为第二十二章所介绍的大量内容提供了信息来源。



## 关于重组 DNA 的辩论

70年代晚期,很多科学家、政治家和有兴趣的非专业人员曾争论重组 DNA 技术的安全性,这个问题首先是由科学家们自己提出的。不久人们认识到,创造显然有害的有机体是不值得的。例如, $\beta$ -内酰胺酶基因被引入到链球菌 (*Streptococcus pyogenes*),引起机体产生“链霉”喉及风湿热;如果这种杂交细菌从实验室逸出进入自然界,那么用来治疗链球菌感染的青霉素就将失效。人们对于从肉毒杆菌中得到的肉毒素基因引入到大肠杆菌中的可能性作出更强烈的警告,因为如果这种细菌菌株建立到大肠杆菌的天然所在地——人的大肠中就可能让肉毒素引起很高的致死态感染,一天之内便死亡。国际上一致同意禁止组建这些显然有危险性的细菌。辩论仍在继续着,然而争论集中在组建的带有外来基因的细菌株是否有毒性上,人们相信当释放到自然界时不至于引起明显的毒害作用。有人认为,科学家没有办法去预言真核基因将是怎样地在大肠杆菌中得到表达。讨论的各种意见开始在展开,关于这方面可见下面的一个例子。假若在一含有人们所感兴趣的基因的 DNA 中可能还含有另一个基因,这个基因在正常情况下不在动物的细胞中表达,但是在细菌中则表达。人可能被这种重组细菌感染,如果这种基因产物分泌的话,它可能引起疾病。因为还没有足够的证据表明这样一种方案可能有害,或可能无害。因此,两种看法相持着,有些人希望看到所有的重组 DNA 研究都被禁止,另一些人则认为如果使用适当的测量可以保证安全。现在没有任何集团否认重组 DNA 研究对科学和人类带来的巨大利益,但是,老的研究组织则坚决主张不能因这种利益而为其危险性作辩护。

争论的结果是美国政府出面制定了一个近乎干预的书面资料,称之为国家健康研究院对于重组 DNA 分子的研究的规则,即人们通常所知的 NIH 准则。这个准则规定了防止实验室中的生物逃逸到环境中去的物理性遏制 (physical containment) 的方法,包括各种隔离工作区以及污物处理。生物学的遏制涉及到宿主细胞方面应当创建一些有生物学缺陷的细菌,这种细菌生存所需要的条件在自然界中不存在。为符合这种要求,进行遗传学设计,创造出一些宿主细胞,例如大肠杆菌  $\lambda 1776$  (它创建于美国二百周年)。这种细菌必须有好几种特殊营养物质作为其生活的底物,这些物质在污水中、土壤中或人的内脏中并不常见,这种细菌通常被污水中的离子去污剂杀死,或被存在于人消化道中的胆汁酸盐杀死,它不能在  $37^{\circ}\text{C}$  (人的体温) 以上生长。因此,如果没有发生好几次的返祖遗传过程,那么这种  $\lambda 1776$  菌就绝不可能顺利地生长在人体之中。后来人们还使用了一些其他的缺陷型宿主以及缺陷型的噬菌体载体。所有使用的质粒载体都不能自我转移,因此,它们自己不能转移到非缺陷型的细菌中。大部分的预防措施已为研究工作者接受 (但是某些科研小组对这些条款并没有全都愉快地接受)。

随着时间的推移以及获得更多的知识,以上的顾虑变得少了,NIH 对某些方面的限制也逐步地放松了。关于对健康有毒害的假设现在似乎已经不那么具有权威性了,虽然这些假设仍然有着社会性的及政治性的价值。到现在为止,还没有报道过重组 DNA 研究产生了不利影响,也没有人报道过由此而引起的什么启示或暗示。相反,此项新技术潜在的巨大利益已被研究者们甚至外行们所认识。但是,又出现了许多新的问题,例如遗传工程已经有工业上的重要性,并成为产生经济效益的一个重要分支。这些新的问题是:第

一, 公司应当得到在他们的实验室创造新的生产形式的专利权吗? 美国最高法院已经肯定了他们的这种权利。第二, 这种经济效益激发人们对未来的研究方向作出新的规划, 这种新的规划应当允许达到怎样的规模呢? 即政府是否应当支持“纯”学术的研究, 而把有经济效益目标的项目留给公司? 第三, 工业竞争是否会使得科学研究的新信息和新思路的相互交流减少? 第四, 得不到经济效益的研究所和大学怎样才能通过竞争而得到一些年青的、最优秀的研究工作者, 因为很普遍的情况是, 研究所在经济上是很困难的, 而较富有的公司则为年青研究者提供高薪。对于这样一类问题目前尚无法简单地作出回答, 人们正在广泛地争论着。

## 参 考 文 献

- Abelson, J., and E. Butz. 1980. "Recombinant DNA." *Science*, 209, 1317—1438.
- Anderson, W. F., and E. G. Diacumakos. 1981. "Genetic engineering in mammalian cells." *Scienc. Amer.*, July, pp. 106—121.
- Beers, R. F., and E. G. Bassett (eds.) 1977. *Recombinant Molecules: Impact on Science and Society*. Raven.
- Brown, D. D. 1973. "The isolation of genes." *Scienc. Amer.*, August, pp. 20—29.
- Cohen, S. N. 1975. "The manipulation of genes." *Scienc. Amer.*, July, pp. 24—33.
- Curtiss, R. 1976. "Genetic manipulation of microorganisms: potential benefits and hazards." *Ann. Rev. Microbiol.*, 30, 507—533.
- Gilbert, W., and L. Villa-Komaroff. 1980. "Useful proteins from recombinant bacteria." *Scienc. Amer.*, April, pp. 74—94.
- Hofschneider, P. H., and W. Goebel. 1982. "Gene cloning in organisms other than *E. coli*." In Arber, W., et al. (eds.) *Current Topics in Microbiology and Immunology*, vol. 96. Springer-Verlag.
- Lobban, P., and A. D. Kaiser. 1973. "Enzymatic end-to-end joining of DNA molecules." *J. Mol Biol.*, 78, 453—471.
- Maniatis, T., R. C. Hardison, E. Lacy, J. Lauer, C. O. Connell, D. Quon, G. K. Sim, and A. Efstratiadis. 1978. "The isolation of structural genes from libraries of eucaryotic DNA." *Cell*, 15, 687—701.
- Mertz, J., and R. Davis. 1972. "Cleavage of DNA: RI restriction enzyme generates cohesive ends." *Proc. Nat. Acad. Sci.*, 69, 3370—3374.
- Public Health Service, U. S. Department of Health, Education, and Welfare. 1976. *National Institutes of Health Guidelines for Research Involving Recombinant DNA Molecules*. U. S. Government Printing Office.
- Scott, W. A., and R. Werner. 1977. *Molecular Cloning of Recombinant DNA*. Academic Press.
- Seeberg, P. H., J. Shine, J. A. Martial, R. D. Ivarie, A. Morris, A. Ullrich, J. D. Baxter, and H. M. Goodman. 1978. "Synthesis of growth hormone by bacteria." *Nature*, 276, 795—798.
- Setlow, J., and A. Hollaender (eds.) 1979—1981. *Genetic Engineering: Principles and Methods*. Vols. 1—3. Plenum.
- Sinsheimer, R. L. 1977. "Recombinant DNA." *Ann. Rev. Biochem.*, 46, 415—438.
- Smith, D. H.: 1979. "Nucleotide sequence specificity of restriction endonucleases." *Science*, 205, 455—462.
- Wu, R. (ed.) 1979. *Recombinant DNA. Methods Enzymol.*, Vol. 68, Academic Press.



## 第二十一章 真核生物病毒

已知有数百种病毒，它们每一种都感染一类特定的动物或植物细胞。许多病毒与严重的疾病有关，但也有些仅仅是细胞内温和的寄生物。起初对病毒进行研究，只是作为一种研究疾病的手段，但很快清楚了各种病毒本身具有各种各样有趣的生化特性，就像研究噬菌体得到了副产品，从而可以解释细菌的许多生化过程一样，对病毒的研究也为进入真核细胞打开了一扇窗户。在研究病毒时，首先观察到它们与真核细胞有一些共同的现象，如基因剪接和多聚蛋白。此外，还发现了一些仅仅病毒才有的独特的生化过程，如逆转录作用。

病毒繁殖方式的多样性，显然比噬菌体的更多些，学生们可能对此感到诧异。然而，这种差别可能只是一种错觉，因为，所有的寄生生物必须适应于其宿主的生化过程；我们所研究过的病毒是在几百种不同宿主中进行繁殖的，而所研究的噬菌体却几乎只局限于三种宿主细菌。

### 真核生物病毒的基本结构

病毒有四种状态：裸露的 20 面体、裸露的螺旋体、有被膜的 20 面体和有被膜的双螺旋体（图 21-1）。每种形态至少有一层外壳包裹着核酸，完整的病毒颗粒称作病毒颗粒（virion）。一种特定的病毒一般只包含一种核酸，它可能是单链或双链的 RNA，也可能是单链或双链的 DNA（见表 21-1）。通常病毒颗粒只包含一个核酸分子，但也有些病毒颗粒有几个分子，其中每个分子一般含有一种基因，偶尔也含有两种基因。

裸露病毒核酸分子的编码容量很小。裸露病毒的颗粒外壳包括一种（植物病毒）或二

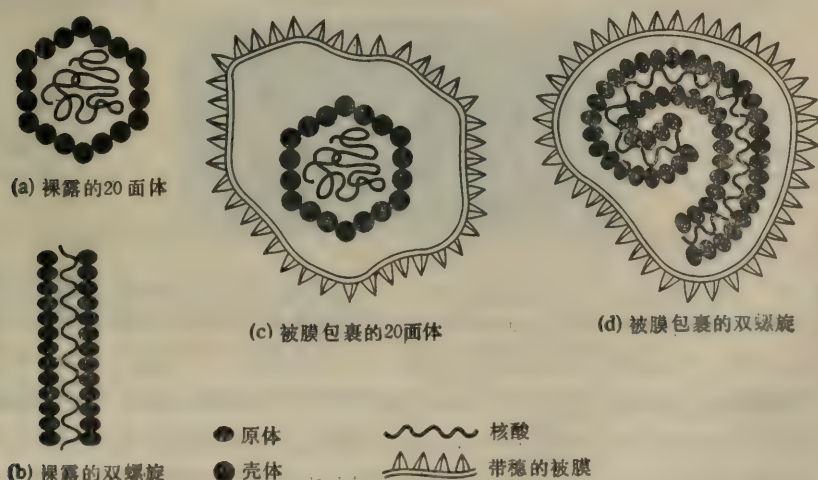


图 21-1 病毒的四种基本形态。

表 21-1 几种病毒和一种类病毒的核酸

核酸形式	代 表 病 毒	分子量 $\times 10^6$	大约基因数
单链 DNA	腺关联病毒 (Adeno-associated virus)	1.5	4—5
	微小病毒 (Parvovirus)	1.5	4—5
双链 DNA	多瘤病毒 (Polyoma virus, SV40)	3	4—5
	乳头瘤病毒 (Papilloma virus)	6	9
单链 RNA 1—2	腺病毒 2 (Adenovirus 2)	23	30
	单纯疱疹病毒 (Herpes simplex virus)	100	150
	牛痘病毒 (Vaccinia virus)	160	240
	马铃薯纺锤类病毒* (Potato spindle viroid)	0.12	无(?)
	烟草随体坏死病毒 (Satellite tobacco necrosis virus)	0.4	1—2
	烟草花叶病毒 (Tobacco mosaic virus)	2	6
	脊髓灰质炎病毒 (Polio virus)	2.5	7—8
双链 RNA	流感病毒 (Influenza virus)	4**	12
	呼肠病毒 (Reovirus)	15**	22
	水稻矮小病毒 (Rice dwarf virus)	15	22

\* 类病毒是有传染性的、无被膜的 RNA 分子,仅存在于受感染的植物中。

\*\* 这些病毒粒子包含几个 RNA 片段。给出的是分子量的总和,每一个片段的分子量范围为  $0.03 \times 10^6 - 3.4 \times 10^6$ 。

种(动物病毒)结构蛋白。组成结构蛋白的构建材料称为原体 (protomer)。偶尔在外壳中也包含有个别的酶分子。

在一裸露的螺旋病毒中,原体环绕着一个单链 RNA 分子呈螺旋状排列。研究得最为详尽的螺旋病毒是烟草花叶病毒,其电镜照片见图 21-2。



(a)



(b)

图 21-2 烟草花叶病毒 (TMV) 的电镜照片。(a) 一完整颗粒, 示出蛋白质和中空核心的盘状排列。(b) 病毒颗粒已破碎, 底片背景上分布着一端向前的片段, 清楚地显示出病毒的中空核心,(b) 放大率约是 (a) 的 1/3 (承 Robley Williams 提供)。

由于病毒是人类疾病和植物疾病的媒介, 因此广泛地研究了裸露的 20 面体病毒。在这些病毒粒子中, 原体组成为壳体 (capsomer) 的集合体。壳体包含 5 个或 6 个原体, 即 5 体 (pentons) 或 6 体 (hexons)。在动物的 20 面体病毒中用两种不同的多肽链形成 5 体和 6 体。但在植物病毒(及 RNA 噬菌体)中则用同种蛋白质, 形成 5 体和 6 体。壳体



以几种方式形成 20 面体的外壳 (capsid)，外壳围绕着核酸，这个“外壳-核酸”单位称作一个核壳 (nueleocapsid)。如果没有被膜，核壳本身就是病毒颗粒(见图 21-3)。

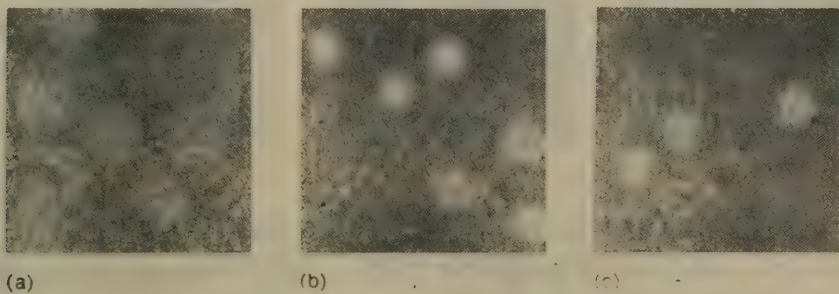


图 21-3 几种不同的 20 面体病毒颗粒的电镜照片。(a) 人类疣病毒 (Humam wart virus)。(b) 伤瘤病毒 (wound-tumor virus)。(c) 蕃茄浓密矮小病毒 (tomato bushy stunt virus) [承 Edward kellenberger 提供 (a), Robley Williams 提供 (b)、(c)]。

表 21-2 病毒颗粒中的一些酶

酶	病 毒	功 能
逆转录酶	逆转录病毒 (Retrovirus)	从单链 RNA 反转录出 DNA; 将单链 DNA 转变成双链 DNA; 核糖核酸酶 H
双链 RNA 转录酶	呼肠病毒 (Reovirus)	从双链 RNA 转录出单链 RNA
DNA 转录酶	痘病病毒 (Poxvirus)	从 DNA 转录出 mRNA
单链 RNA 转录酶	粘液病毒, 杆状病毒, 类粘液病毒 (Myxovirus, rhabdovirus, paramyxovirus)	从单链 RNA 转录出单链 RNA
DNA 核酸外切酶和内切酶	痘病病毒, 逆转录病毒 (Poxvirus, retrovirus)	将 DNA 降解为寡核苷酸、及单核苷酸
三磷酸核苷磷酸转移酶,	很多有被膜的病毒	交换磷酸基
腺 DNA 核酸内切酶	腺病毒 (Adenovirus)	使双链 DNA 产生切口
神经氨酸苷酶或唾液酸苷酶	粘液病毒, 类粘液病毒	切割细胞表面的多糖
蛋白激酶	逆转录病毒, 粘液病毒, 类粘液病毒, 疱疹病毒 (Herpesvirus)	磷酸化蛋白质
tRNA 酰化氨基酸水解酶	很多逆转录病毒	使 tRNA 带电荷
DNA 连接酶	很多逆转录病毒	鐳封单链 DNA 中的缺口

在部分纯化的病毒样品中,经常发现空外壳,说明核酸对外壳的装配并不是必需的。空外壳的结构和核壳的结构略有不同。

有被膜的病毒颗粒并不具有特征结构。在一个有被膜的螺旋病毒颗粒中,螺旋的或 20 面体的核外壳被包裹在一层疏松的、大致呈球形的被膜中。病毒颗粒的形状是可变的,因为被膜并不坚硬。被膜来源于细胞膜。图 21-4 示出了几种有被膜病毒的例子。很多病毒颗粒具有宿主细胞中不存在的编码病毒的酶(表 21-2),在某些情况下,这些酶或是核酸复制酶和转录酶,或是病毒颗粒吸着到宿主细胞上所必需的因子。但在很多情况下,这些酶的确切的功能尚不清楚。





还有未观察到的特殊形式,例如,痘病毒 DNA 是两端封闭的双链分子(见图 12-5)。

倒转重复的碱基序列是线性 DNA 分子的基本特征(图 21-6)。例如,腺病毒的所有单链和双链 DNA 分子都在末端有倒转重复序列,我们以后将会看到这些重复序列对复制来说是至关重要的。疱疹病毒中的末端倒转序列很特殊,即任何疱疹病毒群中的 DNA 分子总是以四种形式存在,它们的 DNA 由两个片段组成,每个片段都可有两种方向,见图 21-6(b)。

病毒的 RNA

图 21-7 示出了几种已知的病毒 RNA 形式,最普遍的形式是单链 RNA。在少数病毒中,例如新培斯 (Sindbis) 病毒,它的倒转末端的重复序列可以形成一个在复制中起作用的双链区域,像单链 DNA 病毒那样。许多哺乳动物的逆病毒是双倍体,而且它们除了 5' 端以外,具有两个相同的 RNA 分子。这两条 RNA 链通过 5' 端氢键相联。具有双倍体 RNA 的逆病毒还含有两个宿主的 tRNA 分子,这两个 tRNA 分子通过氢键连系在病毒 RNA 上,在复制中作为引物起作用。

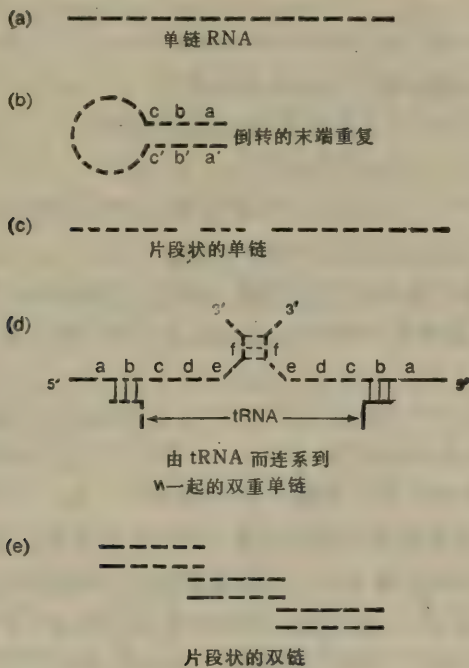
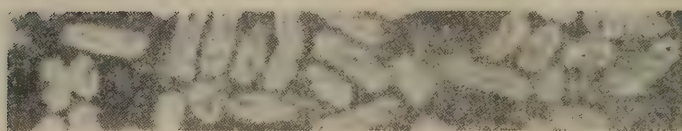
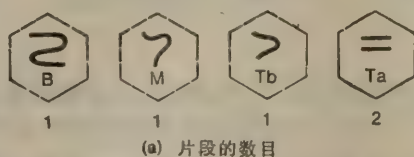


图 21-7 病毒 RNA 的各种形式。(b) 图中环与茎的比例比实际的大得多。注意(d)图中特殊的平行氢键结构。(e) 在某些情况下,片段具有共同的序列,有时这种片段是正常序列。

有些单链 RNA 病毒含有几个不相同的单链 RNA 分子,它们各自携带不同的遗传信息(流感病毒就是一例)。据说这些病毒具有分段基因组(亦称节片状基因组)。苜蓿花叶病毒是一个特别有意思的例子,它带有分段基因组的例子:它的四个 RNA 片段装在四个不同的病毒颗粒中,每个病毒颗粒含有一定量 RNA。苜蓿花叶病毒中的 RNA 种

类分别称做 B、M、Tb 和 Ta RNA,分子量相应为  $11.1$ 、 $0.8$ 、 $0.7$  和  $0.3 \times 10^6$ 。这些片段的包装如图 21-8 所示。成功的感染则需要每种形式的 RNA 各有一个分子进入宿主细胞。带有分段基因组的病毒,如果它的分段基因组的成分包装进不同的病毒颗粒,这种病毒被称为异质外壳病毒 (heterocapsidic viruses); 如果它的所有分段基因组的成分都在同一毒粒中,则这种病毒就称为同质外壳病毒 (isocapsidic)。目前,已知有大约 20 种不同的异质外壳病毒。



·(b)

图 21-8 (a) 苜蓿花叶病毒——一个多成分病毒的四种病毒体。(b) 电镜照片显示含有不同核酸成分的颗粒,宽度相同,但长度不同(承 Edward Kellenberger 提供)。

含有分段基因组的植物蛋白称做多成分病毒 (multiple viruses) 或共病毒 (coviruses)。这种名称不用于有分段的基因组的动物病毒。

病毒中的单链 RNA 分子有两种类型: (+)链和(-)链。

如果一个病毒 RNA 分子具有与病毒 mRNA 相同的极性,并含有可以翻译出病毒蛋白质的编码序列,我们就说它是(+)链 RNA。

一个负链 RNA 分子则是一非编码链,因此它必须再由依赖于 RNA 的 RNA 聚合酶拷贝一次,才产生可翻译的 mRNA。含有以上不同 RNA 的病毒必然有不同的生活周期。

通常,每个病毒粒子(+)链或病毒 mRNA 分子的 5' 端都有一个“帽子”,类似于真核细胞 mRNA 分子中发现的那样(事实上,“帽化”首先是在病毒 RNA 中发现的)。已知唯一的一个例外是脊髓灰质炎病毒,它没有被帽化。在一些具有分段基因组的病毒中,每个片段都有帽子。虽然那些以双链病毒 mRNA 为模板所形成的 mRNA 分子都带有帽子,而作为模板的双链病毒 RNA 分子本身的 5' 末端却是没有帽子的。并且,双链 RNA 病毒(如呼肠病毒)的病毒 RNA 及由它复制出的 mRNA 都没有 polyA 尾。

单链动物病毒 RNA 分子的 3' 末端有一 polyA 尾,通常它比动物细胞中所发现的 mRNA 尾短一些。许多单链的植物病毒 RNA 分子(如芜菁黄色花叶病毒以及雀麦草花叶病毒)的 3' 末端都很特殊,这种 3' 端的碱基序列类似于 tRNA。实际上,在体外实验中,这些分子可从 tRNA 氨基酰合成酶上接受一个氨基酸。然而,体内还没有观察到氨基酰化作用。病毒 RNA 分子末端的性质总结如表 21-3。

一些双链 RNA 的病毒总是具有分段基因组,而且是有同质外壳的。这个类型病毒中被研究得最出色的是有 12 个片段的伤瘤病毒——一种植物病毒,和有 10 个片段的呼肠病毒——一种动物病毒。呼肠病毒的生活周期将在后面讨论。



表 21-3 病毒 RNA 分子的末端性质

RNA 类型	5' 端帽子	3' 端 poly A	相似于 tRNA 的 3' 端
动物病毒			
单链的	通常有	有	无
分段单链的	有, 每个片段均有	有, 每个片段都有	无
双链的	无	无	无
从双链 RNA 拷贝出的 mRNA	有	有	无
植物病毒			
单链的	无	无	很多种病毒都有

这部分指出了 RNA 病毒中存在着多种多样的核酸, 我们将会看到不同的结构需要有产生 mRNA 和蛋白质的不同途径, 以及对这些途径的不同的调控机理。所有的真核 mRNA 分子都是单顺反子 mRNA, 这个性质也导致了核酸复制和蛋白质加工的要求与原核生物的要求有所不同。

## 烈性病毒的基本生活周期

所有烈性病毒的生活周期都包括以下几个过程:

1. 吸附。
2. 核酸进入细胞。
3. 转录、翻译和复制。
4. 颗粒成熟。
5. 颗粒释放。

典型的生活周期为 6—48 小时 (噬菌体只有 20—60 分钟), 每个细胞世代释放出  $10^3$ — $10^5$  个病毒颗粒。后面将要讲到的肿瘤病毒, 则有不同的生活周期。多种病毒的生活周期是如此不同, 以致没有简单的共同机理能够说明上述任一阶段。不过, 我们还是可以把这些阶段的一些主要特点和几个共同的机理作一简单地描述。有关转录、翻译和复制的机理将在本章的其他部分进行详细讨论。

### 病毒吸附寄主细胞

病毒吸附有很多方式, 本书不打算一一赘述。由于细胞表面受体分离的困难, 吸附的研究一直受到阻碍。很可能所有动物病毒的受体都是表面糖蛋白。动物病毒的吸附位点, 对于有被膜病毒来说是一种由蛋白质所构成的短突穗, 对于裸露病毒来说, 或是突穗 (spike), 或是弹性毛 (flexible hair)。目前, 对植物病毒的吸附则知之尚少。

### 病毒核酸进入宿主细胞

病毒核酸转运进入宿主细胞的方式的多样性, 可以和病毒的数量相比。有些病毒只把核酸注入细胞, 另外一些则携带着自身的关键性的聚合酶, 也就是说核酸和蛋白质分子 (或许是核蛋白) 都被转运和穿入细胞膜。通过对少数几种有被膜病毒的研究, 得到的一些证据表明, 穿入的早期阶段是病毒的被膜与细胞的细胞膜融合, 然后细胞膜的成分就包围了核壳。无论是有被膜的还是裸露的病毒, 紧跟早期阶段的下一步都是构成外壳的蛋白

质的水解消化过程，称作脱壳（uncoating）。早先，有些教科书介绍过几种脱壳机理的详细模型，现在看来，这些模型都不正确。目前我们尚未真正了解穿入的全部过程，但是很可能这个过程一直受着被膜中或外壳中病毒编码的蛋白和特殊的膜蛋白之间的专一性相互作用的调节。

### 子代病毒颗粒的成熟和释放

与大多数烈性噬菌体的生活周期不同，烈性动植物病毒的生活周期并不总是以细胞的逐渐死亡和裂解为终点。由于受到某些病毒的侵染，一个被感染的细胞通常会变得不正常，生长缓慢，可能最终会死去；但同样也普遍地存在着这种情况：被某种病毒侵染的细胞会无限地生产，不断释放出子代病毒（受感染的细胞经常合成和释放特异蛋白质。例如释放出形成腺病毒外壳的5体，此物质可能与受感染的有机体对病毒的致病作出反应有关）。

病毒颗粒由病毒核酸和病毒蛋白质装配而成并大量地产生着。烟草花叶病毒的组装模式已在第六章详细描述过了，在这一部分，我们将要对有些病毒的成熟和释放的偶联的过程加以介绍。

病毒释放有三个基本历程：

1. 病毒从细胞膜下所形成的空泡（vacuoles）中释放出来。
2. 裂解。
3. 出芽。

前两条是裸露病毒的特征，而出芽是有被膜病毒的特征。每个历程都显示出很多变化。这些差别在某种程度上是由于不同病毒组装的位置不同造成的——有些病毒进行复制，并在细胞质中组装；而另一些病毒则成为核病毒。细胞质中的病毒颗粒常常恰好堆积在位于细胞膜下的空泡中，随着空泡的破裂释放出来（图 21-9），但并不破坏完整的细胞。这一历程通常在病毒感染后 1—2 天才发生。但有些病毒如已知的快速释放病毒，在感染后 6—8 小时就脱整合，释放出病毒颗粒。核病毒通常不通过形成空泡而释放，而是在核内，有时也在细胞质中，形成大量结晶体排列（guasicry stalline）而排出。大约在感染 24—48 小时后，细胞开始脱整合，释放出病毒颗粒，有些在细胞质中倍增的病毒也通过此方式裂解细胞。有被膜病毒通过出芽（budding）而释放（图 21-10）。在这种释放方式下，细胞质中合成的病毒编码蛋白掺入到细胞膜中，引起了膜结构的重新组建（reorganization）。核外壳粘连到细胞膜中的这些新的病毒蛋白上，使膜在核外壳外周形成一个突起球面。通过一种还不知道的方式（可能是穿透过程中使用的步骤），细胞膜在芽泡之下重新形成，释放出包在被膜内的核壳颗粒。因此认为被膜是细胞膜的一部分。

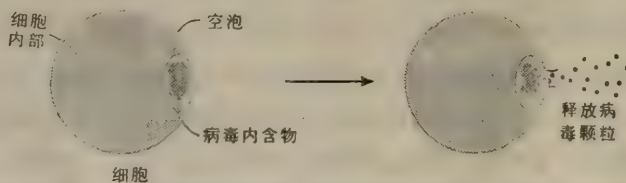


图 21-9 通过空泡形成释放病毒颗粒。一个病毒颗粒聚合物——病毒内含物在细胞膜下的一个空泡形成。



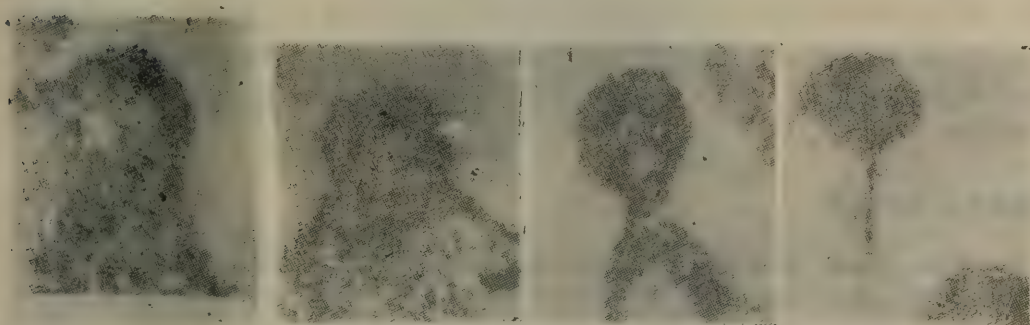
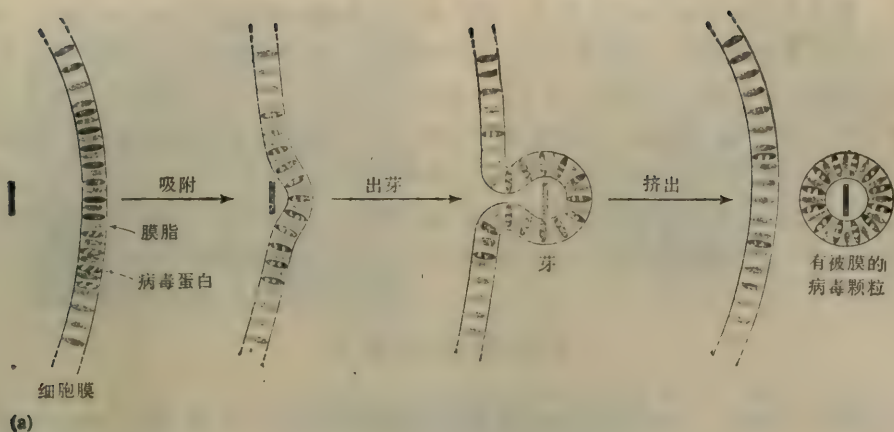


图 21-10 (a) 通过出芽方式形成有被膜的螺旋蛋白。实际上,膜比病毒粒子的宽度薄得多。脂和病毒蛋白也并非如图所示那样排列有序。(b) 出芽过程的电镜照片 (承Edward Kellenberger 提供)。

分, 细胞膜中含有病毒的蛋白质而没有细胞质中的蛋白质。

### 宿主细胞被病毒接管

在我们讨论噬菌体生活周期时, 有一个阶段定义为噬菌体“接管” (takeover) 细菌宿主。在这个时期, 细菌被转化成为制造噬菌体的机器, 这种接管作用的一部分是指在噬菌体生活周期的早期, 停止或减少宿主 DNA、RNA 及蛋白质的合成。人们注意到典型的动物病毒生活周期中缺少这一阶段。这种典型阶段经过验证, 确实存在于某些病毒生活周期中, 但当然不是普遍的。实际上, 病毒和宿主常常共生一段时间, 只是在感染晚期宿主的功能才被破坏。

有些病毒的生活周期中有特定的时期, 在这些时期中宿主受到剧烈地改造, 但还不清楚这种改造对于病毒的最适产生是否必需, 在少数体系中已经发现有以下三种不同形式:

1. 有时, 宿主 RNA 和 DNA 的合成一直到病毒的生活周期的很晚期才受到抑制, 这种抑制很可能是在病毒核酸的复制作用开始之后才起作用。

2. 有时, 核糖体的合成受到削弱; rRNA 基因的转录是正常的, 但 rRNA 前体的加工受到抑制。

3. 细胞质蛋白质的合成受到抑制, 这是一种普遍现象, 虽然抑制的程度和时间因不同病毒而有很大变化。脊髓灰质炎病毒和腺病毒的这一过程是已知的, 在脊髓灰质炎病毒的

感染中, 宿主多肽链的起始受到抑制。宿主蛋白质合成的起始复合物 mRNA-核糖体-tRNA<sup>met</sup> 的形成有缺陷。在病毒 RNA 结构那部分已提到脊髓灰质炎病毒颗粒 RNA 是 (+) 链的缺少帽子的核糖核酸。显然, 病毒蛋白质(目前尚未分离到这种蛋白质)对于 5' 端有帽子的 RNA 翻译成蛋白质的过程有影响, 它干扰翻译起始物的形成, 但是对无帽 RNA 的翻译不干扰。对于腺病毒来说, 宿主的细胞 mRNA 是在核中制造的, 这时运来被加工, 病毒蛋白质的干扰就阻止了它向细胞质的转运, 因此这种 mRNA 得不到翻译。而那些在腺病毒感染前就已存在的宿主细胞质中的长寿的 mRNA 也不能翻译。

## 病毒颗粒的鉴定

无论作什么实验, 我们都要知道病毒和细胞的比率——感染倍数 (MOI, multiplicity of infection)。病毒可以通过电子显微镜进行计数, 但是这个方法有以下两个缺点: (1) 由于技术上的原因测量值可以偏大或偏小 3 倍之多; (2) 没有办法知道哪部分颗粒是有感染性的。因此, 发展了类似于滴定噬菌体形成噬菌斑的鉴定计数方法。下面就介绍其中的一些方法。

### 病毒的生物学鉴定

许多动物细胞可在培养皿中生长成为单层细胞。当生长完成时, 可被各种染料着色的细胞的融合形成一个有色的细胞层。往培养皿中先加入细胞, 紧接着加病毒, 于是其中一些病毒将会感染生长层中的细胞。如果是致死型的病毒(如脊髓灰质炎病毒), 或者至少是对细胞生长有显著抑制作用的病毒, 那么当单层细胞停止生长时, 在这一单层中就会出现一个区域, 其中极少或甚至没有细胞。因此, 在染色背景上将出现无色的环状区域——噬斑 (plaque) (图 21-11(a))。与噬菌体形成噬斑只需要 3—12 小时相比, 后者需要 3—21 天, 因为动物细胞生长缓慢, 不同的病毒生成噬斑的天数不同。

有些病毒并不形成噬斑, 因为它们不能尽快地杀死细胞。但是, 如果一种病毒(如流感病毒)能刺激红细胞凝集素 (hemagglutinin) 的合成, 那么也有办法鉴定感染颗粒。凝集素是一种糖蛋白, 在感染细胞的表面形成。它使感染细胞与红细胞结合。因此, 如果细胞的融合层中含有的感染集中区病灶不像噬斑那样容易看到, 那么, 我们就可以再铺一层红细胞悬浮液进行观测: 铺层, 再清洗, 经过以上处理后, 就可以从单层细胞表面上的红斑点鉴别出这种病灶的存在。

在噬斑鉴定法出现以前, 某些病毒(如牛痘病毒和疱疹病毒)是通过痘疹法 (pock method) 鉴定的。如果使发育 10 天的鸡胚的绒毛膜受病毒感染, 然后再培养 36—72 小时, 在透明的膜上就出现了不透明的集中区, 称之为痘疱。痘疱相当明显, 容易计数。

肿瘤病毒(例如劳氏肉瘤病毒)也可以在融合细胞上计数。这些病毒可以使受感染的细胞在融合层停止生长之后, 还继续生长好几天, 形成称之为病灶的生长区, 它在单层细胞上好像突起的一团, 见图 21-11(b)。

植物病毒(如烟草花叶病毒)的计数方法如下: 刮下一层叶表面, 在其上铺上一滴病毒悬浮液, 于是叶子上就形成了死斑(又称为侵蚀斑, lesions); 每个侵蚀斑都是由一个病毒颗粒侵染所形成的。



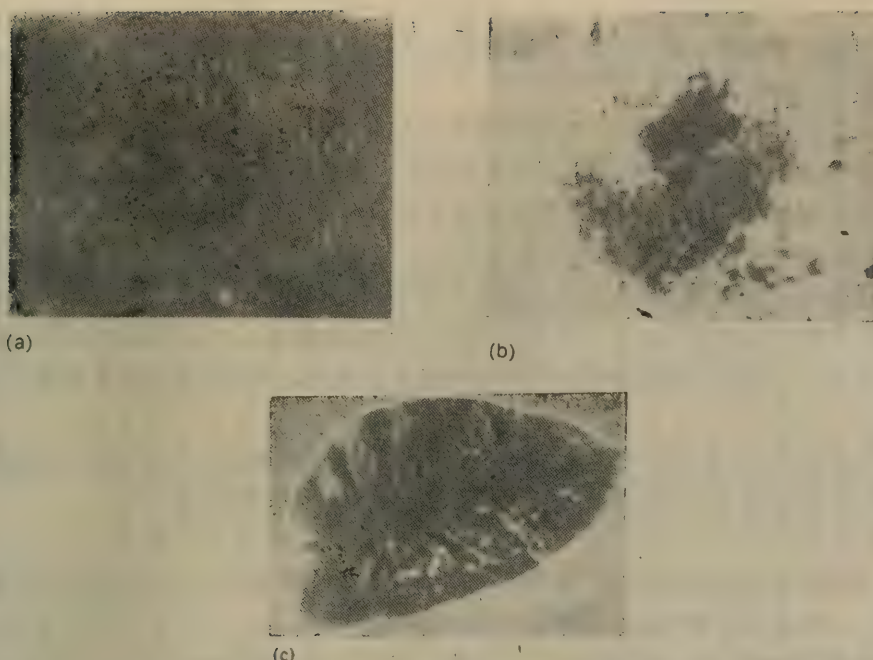


图 21-11 检测病毒的三种生物学方法。(a) 腺病毒在染色的动物细胞的融合背景上形成噬胞斑。(b) 肿瘤病毒的病灶。照片显示劳氏肉瘤病毒的病灶，其中的黑色区域是一团受病毒感染的细胞所形成的细胞簇，它们是生长于未受感染的单层鸡胚细胞上的多层细胞。(c) 受烟草花叶病毒感染烟草叶子，照片显示了干性组织的侵蚀斑。

## 检测病毒的效率

上述各种生物学鉴定病毒的方法效率都很低，而且随病毒的不同而变化（见表 21-4）。与在电镜下所见的颗粒数相比较，只有 0.01—10% 的动物病毒颗粒产生可见到的感染（塞姆利基氏森林病毒和单纯性疱疹病毒，则是值得注意的例外的情况。在最适条件下进行鉴定，最好的制剂可以给出 100% 感染效率）。植物病毒的效率更低，如表中所列，低效感染的原因可能有以下三种：

1. 由于装配出有缺陷的颗粒或在装配后受到某种损伤，大多数病毒颗粒是无感染力

表 24-2 病毒颗粒与感染单位之比(用各种测定方法)

病 毒	比 例*
动物病毒	
塞姆利基氏森林病毒 (Semliki forest virus)	1
痘病毒 (Poxvirus)	1—100
流感病毒 (Influenza)	7—10
呼肠病毒 (Reovirus)	
腺病毒 (Adenovirus)	10—20
脊髓灰质炎病毒 (Poliovirus)	30—1000
猿猴多瘤病毒 40 (SV40)	
植物病毒	
烟草花叶病毒 (Tobacco mosaic virus)	$5 \times 10^4 - 10^6$

\* 比例范围很宽，表明因选用的测定方法不同，所以差异较大。

的。

2. 可能由于某些未知的原因,使检测病毒的条件限制了病毒形成噬斑、痘疱、病灶或其他最终斑点的能力。

3. 成功的感染需要许多有功能的病毒颗粒的共同作用。如果有人要了解病毒生活周期的分子学基础,那么很重要的一点就是要确定这一条是否正确,即是否为共同作用。与动物病毒及噬菌体不同,多组分的植物病毒不同颗粒携带着不同 RNA 分子(图 21-8)。此外对动物病毒来说,成功的感染总是只需要一个有功能的病毒颗粒;对噬菌体的感染来说,也是这样。这个看法已由动物病毒形成噬斑(或其他病毒所形成的病灶)的数目与所加入的病毒颗粒的数目直接地成正比的事实证实。如果有人对这个结论的争论有兴趣,他可以按照一个成功的感染需要两个病毒的假说去考虑,那么他就能很容易地作出正确的判断。

按照泊松分布,如果病毒颗粒总数/细胞数为  $m$  (在常规化验中培养皿上的数目是很少的),那么细胞被  $k$  个颗粒感染的比例  $P(k)$  (其实就是噬斑数)为:

$$P(k) = e^{-m} m^k / k!$$

如果认为非要有两个病毒颗粒才能实现成功的感染,于是就可以认为有两种形式的未感染细胞:一种是未受病毒颗粒感染的细胞,其比例  $P(k)$  为  $e^{-m}$ , 另一种是只受一个病毒颗粒感染的细胞,其比例为  $me^{-m}$ , 因此, 不产生噬斑细胞的比例值  $P(0) = e^{-m} + me^{-m}$ , 或  $e^{-m}(1 + m)$ 。在  $m$  很小时,上述表达式可写为:  $P(0) = 1 - \frac{1}{2} m^2$ 。因为受感

染细胞的比例  $P_i = 1 - P(0)$ , 所以  $P_i = \frac{1}{2} m^2$ 。这就意味着噬斑的数目与加入病毒

颗粒数的平方成正比。很明显,如果造成一次成功的感染,所需病毒颗粒数大于 2, 那么,

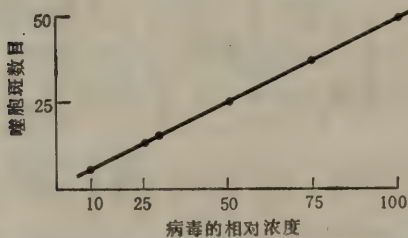


图 21-12 数据表明噬斑数正比于加入的病毒颗粒数。在指定浓度下,向一个接种着有足够细胞的培养皿中加入一定体积的病毒悬浮液形成一融合层。

噬斑的数目将随着加入的病毒颗粒数的平方而增加。如果是另一种情况,即成功的感染过程必需的病毒粒子数只是 1, 则未受感染细胞的部分是  $e^{-m}$ 。受感染细胞的部分是  $1 - e^{-m}$ 。如果加入的病毒颗粒增加  $q$  倍,那么受感染的细胞的部分将是  $1 - e^{-qm}$ 。比值  $(1 - e^{-qm}) / (1 - e^{-m})$  就是当加入  $q$  倍的病毒时,噬斑增加的系数。当  $m$  较小时,此式等于  $q$ 。因此,当加入的病毒增加  $q$  倍时,噬斑的数亦增加  $q$  倍。由于  $m$  值很小,所以此结论

独立于  $m$  值,即此结论不受加入的颗粒数影响。图 21-12 中的线性曲线表明每一个病毒颗粒就可以引起一次感染,从而证实了前面提到的看法。

## 动物病毒的分类

动物病毒将是本章的主要课题,在分子生物学研究中使用了很多动物病毒,并且它们的名称是如此地复杂,因此,列出表 21-5 以供进一步参考。此表列出了动物病毒的主要



表 21-5 分子生物学研究中使用的病毒

病毒家族	名称的来源	核酸的类型	代表性病毒	是否形成肿瘤
腺病毒 (Adenovirus) 疱疹病毒 (Herpesvirus)	腺样组织 Herpein 来自希腊文“爬”	ds DNA (双链) ds DNA	腺病毒 (Adenovirus) 鸡痘病毒 (Herpes zoster)	某些 某些
痘病毒 (Poxvirus)	形成痘疱	ds DNA	单纯疱疹病毒 (Herpes simplex I) 牛痘病毒 (Vaccinia) 纤维瘤病毒 (Fibroma)	某些
小核糖核酸病毒 (Picornavirus)	Pico 即“小”, 加上 rna (即 RNA)	ss RNA (单链)	有一些粘液瘤病毒 (Myxoma) 脊髓灰质炎病毒 (Polio) 埃柯病毒 (ECHO) 门戎病毒 (Mengovirus)	不
粘液病毒 (Myxovirus) 杆状病毒 (Rhabdovirus) 披膜病毒或虫媒病毒 (Togavirus 或 arbovirus)	myx 拉丁文中是粘液 Rhabdos 在希腊文中是“棒”的意思 Toga 在拉丁文中是“覆盖”或“外衣” arbovirus 来源于 arthropod-borne	ss RNA ss RNA (-) ss RNA (+)	流感病毒 (Influenza) 疱疹口炎病毒 (Vesicular stomatitis) 黄热病毒 (Yellow fever) 新塔斯病毒 (Sindbis) 塞姆利基森林病毒 (Semliki forest) 东方马病毒 (Eastern equine) 呼肠病毒 (Reovirus)	不 不 不
呼肠病毒 (Reovirus) 多瘤病毒*(Polyoma)	呼吸道肠道孤儿 拉丁文 Poly 是“许多”, oma 是“肿瘤”	ds RNA ds DNA	多瘤病毒 (Polyoma) 猿猴多瘤病毒 40(SV40)	不是 是
乳多瘤病毒 (Papilloma) 逆转病毒 (Retrovirus)	Papilla 拉丁文的“乳头”或“疣”(或形成疣) Retro 拉丁文为“反,逆”	ds DNA ss RNA	乳头状瘤病毒 (Shope papilloma) 禽类白血病病毒 (Avian leukosis) 鼠白血病 (Murine leukemia) 弗云德白血病病毒 (Friend leukemia) 莫洛伊病毒 (Moloney) 劳氏肉瘤病毒 (Rous sarcoma) 小鼠乳腺瘤病毒 (Mouse mammary tumor)	是 是

\* 乳多瘤病毒 (Papilloma)、多瘤病毒 (Polyoma) 和泡形病毒 (Vacuolating virus) 都属于乳多孔病毒 (Papova virus, 或称空泡型乳头状多瘤病毒) 的家族。

家族、研究中最常见的特殊病毒以及它们名称的来源。

## 动物 RNA 病毒

研究动物 RNA 病毒必须考虑以下几点:

1. mRNA 的产生机理可以用来区分各类 RNA 病毒的基本特征;也就是说,特定的机理决定着复制和翻译的基本形式。

2. 所有的真核 mRNA 分子都是单顺反子,而且真核核糖体很少可能从单个的 mRNA 分子翻译出一个以上的多肽链;然而,所有已知的 RNA 病毒都指导合成几种蛋白质。

3. (-)链 RNA 病毒在翻译前必须合成一个(+)链 RNA。迄今,还未发现宿主细胞中有能够拷贝出 RNA 分子的 RNA 复制酶。

4. 单链病毒颗粒的 RNA 分子不能作为模板直接地复制颗粒的 RNA;因此,为了 RNA 的复制,必须制造出含有互补链的或可组成互补链的中间物。

对于许多 RNA 动物病毒来说,复制作用和产生病毒 mRNA 的过程是互相依赖的,最好是放在一起加以讨论。产生 mRNA 有四种主要的机理,这四种形式还有许多变化形式,图 21-13 描绘了这四种主要机理。在每种类型中可以很方便地挑选一种特定的病毒为例,检查这种病毒的复制、转录和翻译模式。所选择的病毒在图中已标明所选择的病毒的名称。

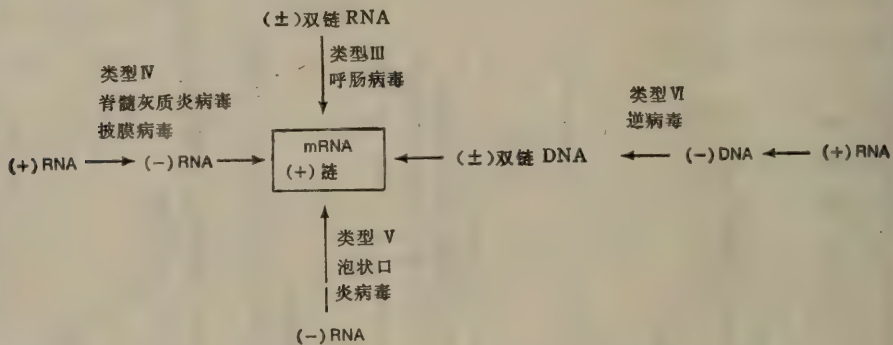


图 21-13 mRNA 产生的四种机理。与这些机理相关,文中讨论的病毒分类也在图中示出。I 型、II 型都是 DNA 病毒,将在别处加以讨论。

### 病毒的 RNA 聚合酶

在我们讨论个别病毒之前,有必要对聚合酶作一些介绍。许多病毒颗粒携带着它们自己的 RNA 聚合酶。对于(-)链 RNA 病毒不能在没有自己的聚合酶的条件下发挥功能,因为动物细胞不含有依赖于 RNA 的 RNA 聚合酶,所以在感染中,这种病毒的感染性(-)链被定义为不可翻译的链。对于双链的 RNA 病毒来说,病毒的聚合酶也是关键的,因为双螺旋中没有可供使用的(+)链去结合到核糖体上并进行翻译,必须任其(-)链再复制出一条新(+)链,才能翻译。而且宿主细胞细胞质中,没有已知的聚合酶能以双链 RNA 为模板合成 RNA。对于(+)链 RNA 病毒,这种病毒对结合到病毒核酸上的 RNA 聚合酶无明显的要求。但是,对于具有 DNA 中间物的逆病毒(DNA 是这



类病毒生活周期的独特特征),则一定需要能结合到病毒 RNA 上的 DNA 聚合酶(即逆转录酶),因为在已知的动物细胞中,从未见到在 RNA 模板上合成出 DNA。

泡状口炎病毒

泡状口炎病毒,即 VSV (vesicular stomatitis virus) 是一种小牛疾病的病原体,它包含一个分子量为  $4.0 \times 10^6$  的 (-) 链 RNA 分子(见图 21-14)。这个(-)链 RNA 分子既无 5'-帽子,也无 3'-polyA 尾,它编码以下五种蛋白质:

- N 主要的病毒外壳蛋白。
- L, NS 发现于外壳中,为 RNA 合成所需要。
- M 被膜蛋白。
- G 穗状蛋白。

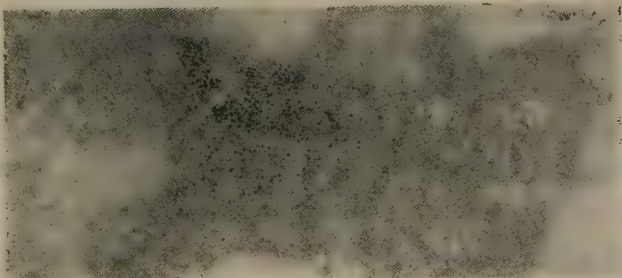


图 21-14 泡状口炎病毒的电镜照片(承 Robley Williams 提供)。

为了合成这五种蛋白, VSV 必须形成(+)链 mRNA, 并必须克服动物细胞不能对多顺反子 mRNA 进行翻译的问题。VSV 解决这些问题的办法是合成五种不同的 mRNA 分子,如图 21-15 所示。这些 mRNA 分子通过 L 和 NS 酶的共同作用而合成, L, NS 酶则被带入细胞中,在 N 蛋白分子所包裹的(-)链 RNA 病毒粒子上。这些 mRNA 的

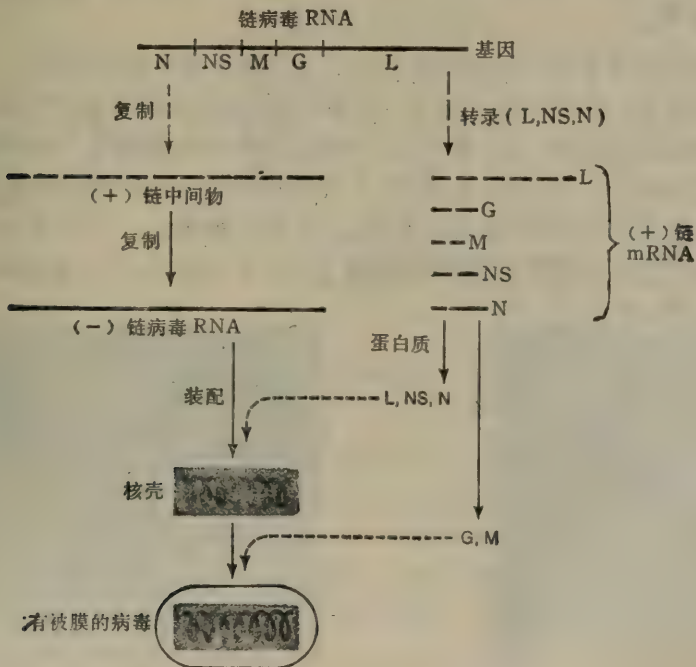


图 21-15 泡状口炎病毒的复制作用与转录作用的图解。

每一个分子都带有 5'-帽子, 并有 3'-polyA 尾, 而且都只包含一个顺反子。还不能确切知道究竟是先合成一个大分子(+)链 RNA, 然后再切成五个小 RNA 分子, 还是通过相继的起始和终止, 先后地合成五个分子 RNA。已有初步的证据表明, 不同的(+)链 RNA 分子是通过一系列连续的起始和终止而依次合成的。

VSV 的复制也是通过先从病毒 mRNA 拷贝出一个(+)链 RNA 而进行的, 这些全长的 RNA 链也是在 L 和 NS 蛋白的作用下形成的。但是, 它们不被切割(它们无 mRNA 功能), 而是立即作为模板, 在 L 和 NS 酶催化下, 复制出一个(-)链病毒 RNA。当(-)链 RNA 形成时, 那些早些时候合成的 mRNA 已经翻译出很多的 N 蛋白了; 于是病毒 RNA 和 N 蛋白迅速地结合, 形成核壳。

被切割的(+)链 mRNA 的转录过程和以(+)链为模板形成(-)链 RNA 的复制过程是以某种方式相互分开的。人们找到了只有其中一个过程受抑制的突变株, 从而得出以上看法。

### 节片状(-)链病毒

关于节片状(-)链病毒(即分段基因组(-)链病毒)(segmented (-) strand viruses)复制和转录的模式, 还不是很清楚。流感病毒即是其中的一个成员。但是, 既然每个片段只包含一个顺反子, 就不需要产生单顺反子 mRNA 分子的特殊步骤。相反, 这些病毒所面对的问题是如何在每个核壳中得到一套完整的片段, 有两种可能性: (1) 有一种特殊机理, 为每一个核壳在每种型式中选出一个分子; (2) 分子装配是随机的, 因此大多数病毒颗粒是不完整的, 是有缺陷的。虽然, 还没有很确定的事实, 然而, 流感病毒的情况似乎很可能是随机装配的。

### 脊髓灰质炎病毒

脊髓灰质炎病毒 (poliovirus)(图 21-16) 是了解得最清楚的病毒之一, 它包含一个分子量为  $2.5 \times 10^6$  的(+)链 RNA 分子。病毒在敏感细胞的细胞质中复制, 形成成熟病毒颗粒的、大的结晶包涵体(图 21-17)。当受损伤的细胞裂解时, 病毒颗粒就释放出来。病毒颗粒 RNA 有一个 3'-poly A 尾, 但 5' 端未被帽子化, 而是共价地连接着一个功能未知的小蛋白质。RNA 分子编码七个蛋白质——四个膜蛋白、一个 RNA 复制酶及两个功能未知的蛋白质。外壳中不携带酶, 病毒 RNA 有一些 mRNA 的活性, 但并不指导所有的蛋白质合成。



图 21-16 脊髓灰质炎病毒的电镜照片  
(承 Robley Williams 提供)。

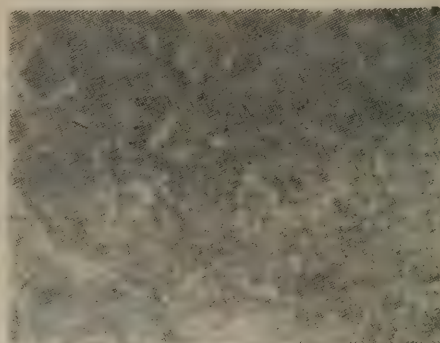


图 21-17 在感染细胞的细胞质中成熟脊髓灰质炎病毒的结晶内含物(承 J. Meyer, M. Weiss 和 R. Böni 提供)。



进入细胞后,病毒 RNA 链立即部分地被翻译,合成出 RNA 复制酶。RNA 复制酶是从一个较大的多肽链经加工而成,这将在下一部分介绍。此酶催化合成出(-)链 RNA (图 21-18)。此(-)链 RNA 唯一的功能显然就是作为进一步合成(+)链 RNA 的模板。从 3' 末端起始,拷贝出(-)链 RNA。注意,第二条(+)链的合成并不是要等到第一条链完成后才开始,而通常是五条(+)链同时以(-)链 RNA 为模板,通过一个多分枝的复制中间物而形成。这些新(+)链起三种作用:(1)用作 mRNA;(2)用作进一步合成(-)链的模板;(3)作为子代病毒颗粒中的病毒 RNA。此时,这些(+)链缺乏 5' 端蛋白。5' 端蛋白是在装配过程中加上去的。

脊髓灰质炎病毒复制过程的四个特征需要进一步评论,下面列出这些特征,并加以讨论:

1. 新生 RNA 并不完全与模板碱基配对。图 21-18 所示分子中,包括一个完整的(+)链和一个不完全的(-)链。有人也许以为新生的 RNA 将与(-)链的互补片段进行碱基配对,形成一个双链区域。虽然还没有得到证实,人们还是认为新生的 RNA 被一种蛋白质所包裹,蛋白质的结合阻碍了这种配对过程。

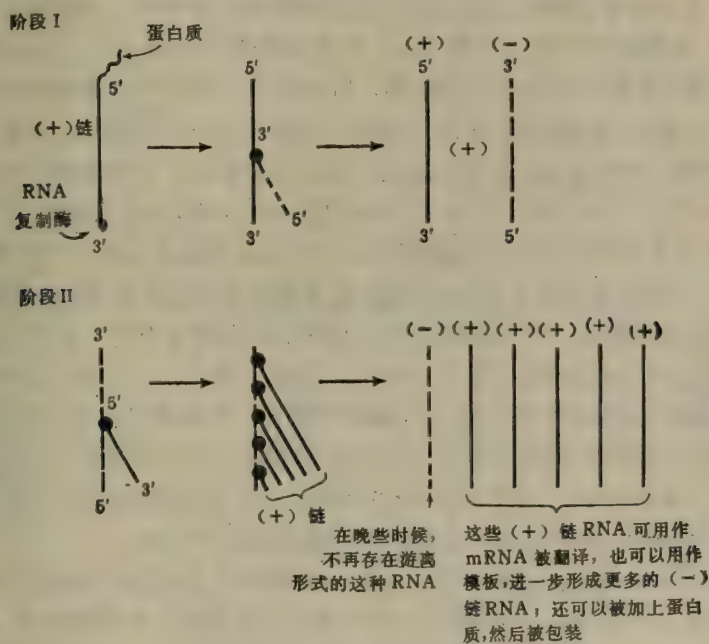


图 21-18 脊髓灰质炎病毒 RNA 的复制。在阶段 I 产生(-)链模板;在阶段 II 产生许多(+)链。

2. 在感染的细胞中发现了双链 RNA。双链 RNA 的存在引出了一种看法:复制的第一阶段是由(+)链病毒 RNA 向双链 RNA 的一种转变。后来,这种看法又因为有几点证据而取消。现在认为,子代链(如图 21-18 中的阶段 II 中的第一个分子)可能由于缺少一种通常抑制碱基配对的 RNA 结合蛋白,因而可与模板通过氢键形成双链 RNA。双链 RNA 在病毒的生活周期中被认为是没有功能的。

3. 可以分离得到复制中间物,它包含着一个长的、有缺刻的、带有单链分枝的双链片段。这种结构只能在 RNA 成分非常纯净、那种阻碍新生 RNA 进行碱基配对的蛋白质

已被除去的制品中才被发现。人们认为这种复制中间物结构也许是由于纯化分离而造成的一种假像。图 21-19 示出了这个机理。

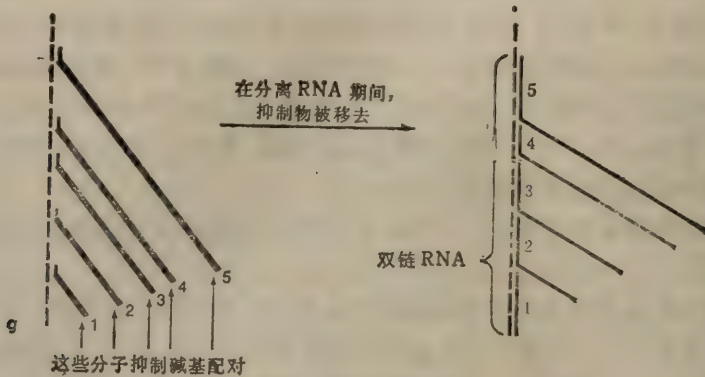


图 21-19 在分离过程中产生复制中间物假像的可能机理。

4. 脊髓灰质炎病毒 RNA 的量先以指数方式增长几个小时, 然后随时间而线性地增加。图 21-18 表明脊髓灰质炎病毒 RNA 的复制按两个阶段进行。在合成(一)链 RNA 的阶段 I 结束时, 有两个分子。在阶段 II, 以阶段 I 合成的(一)链为模板, 合成出许多条子代(+)链。亲代(+)链可以再作为模板, 合成出另一条(一)链。如图所示, 在阶段 II 中, 几条(+)链可同时合成。然而在感染早期因为复制酶分子有缺陷, 所以, 开始时的阶段 II 只形成一个子代(+)链。因此, 在阶段 II 合成的早期, 合成完成时, 产生 4 个 RNA 分子; 当阶段 I 和 II 的循环同时再度重复时, 产生出 8 个分子; 第三次重复产生出 16 个分子; 依次类推。进一步可以了解到, 在每个复制循环之间还有明显的时间延迟, 因为, 每合成一个脊髓灰质炎病毒 RNA 分子只需一分钟, 而上述几个分子复制的表观倍增时间却是 15 分钟。这种倍增过程是指数方式增长的过程, 大约维持 2 小时, 直到有大约 600 个分子为止。在此扩增循环中, 复制酶分子的数目通过反复地翻译和加工而逐渐地加多, 多至像外壳蛋白分子的数目那样。于是, 多分枝的复制方式占了优势。从此(+)链 RNA 分子数随着时间呈线性增长。新产生的(+)链被翻译或者被装配, 而(一)链则只是偶然地产生着。因此在后期主要的合成就是按线性动力学产生(+)链。

脊髓灰质炎病毒 mRNA 编码七种蛋白质。如前所述, 既然真核细胞中一个 mRNA 分子一般只能翻译一个多肽链, 那么要产生这些蛋白质就需要一些特殊机理。从以下两个观察中可以看出, 在脊髓灰质炎病毒感染的细胞中, 蛋白质的合成有以下特殊的特征:

1. 在知道脊髓灰质炎病毒基因数目之前, 曾发现受<sup>35</sup>S-蛋氨酸(它可以标记蛋白质)标记并受这种病毒感染的细胞合成出十种蛋白质, 这些蛋白质的分子量超出了 mRNA 的编码容量。对此现象的可能解释为基因是重叠的(当时这个想法曾被认为是荒谬的), 有些蛋白质是某些蛋白质的切割产物。现在, 这种解释已被人们承认\*。

2. 在<sup>35</sup>S-蛋氨酸的脉冲-追踪实验中, 有些标记的蛋白质在追踪中消失了。与此同时, 放射性标记又出现在那些分子量比最初标记的蛋白质分子量低一些的蛋白质中。这些情

\* 在这些实验完成时, “一个基因一个酶”的假说已被承认为学说。遗传学分析与碱基序列分析对大肠杆菌  $\phi$ X 174 得出清楚的、很能说明问题的数据之后, 人们才开始认真地考虑“基因也许有重叠”的可能性。



况清楚地表明,在脊髓灰质炎病毒蛋白质的形成中有切割过程。

几个脉冲标记实验导致了下面关于脊髓灰质炎病毒蛋白质产生的说明图(图 21-20)。从 mRNA 翻译的单个巨大多肽,称为多聚蛋白(polyprotein)。在多聚蛋白中,当处于蛋白质 NX 两头的两个肽键被切开时,立刻形成两个切割片段 N1 及 NX。因此,蛋白 N1 和蛋白质 NX 在蛋白 N1.5 完全合成之前就从正在生长的多聚蛋白链上释放出来。这些切割片段叫做新生片段,因为它们是从一个生长(新生)多肽链上合成的。一旦 N1.5 释放出来,就再从 N1.5 中移走一个小的 NH<sub>2</sub> 末端片段 N3, N1.5 就转变成 N2, N2 再被切割形成 N4 和 N5。对这些连在一起的 N 蛋白质还未进行很好的研究,但知道至少其中的一些参与 RNA 的合成。RNA 复制酶还未纯化,还不知道复制酶是这些蛋白质中的一个,还是这些蛋白质中的某几个所组成的聚集集体。

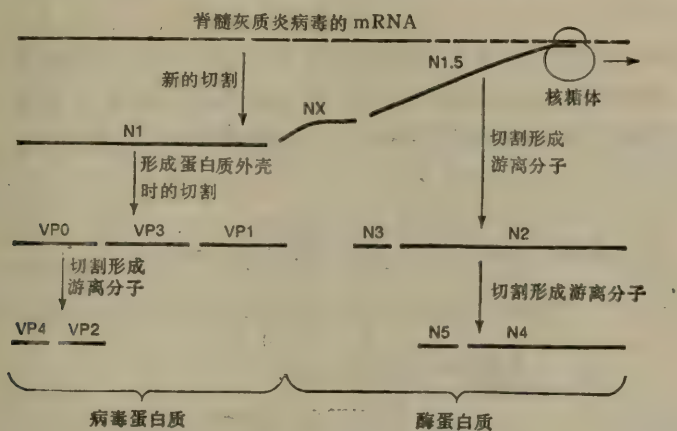


图 21-20 通过连续切割前合成脊髓灰质炎病毒蛋白质。符号 VP 和 N 分别表示病毒的外壳蛋白和非病毒的蛋白质。

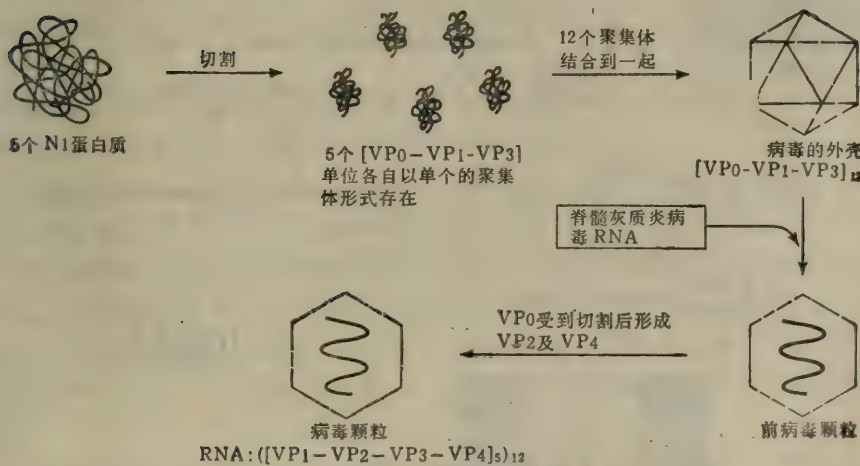


图 21-21 在脊髓灰质炎病毒的形成中蛋白质的加工。

N1 蛋白质在形成病毒外壳过程中将受到一系列的切割,见图 21-21。五个 N1 蛋白分子聚合物形成一个 5 体(pentamer)。将各个 N1 彼此切开,形成五个单位。每个单位(N1)中含有一个 VP0,一个 VP1 和一个 VP3 蛋白,这些蛋白质聚集形成的单

位被称为 5 体原 (propenton)。十二个这种 5 体原进一步聚合,形成外壳原(procapsid)。这是一个大小和形状相同于病毒颗粒的空粒子。当脊髓灰质炎病毒的(+)链 RNA 进入外壳原后,形成一个病毒原。最后,当 VP0 被切割形成 VP2 和 VP4, 四种蛋白质 VP1, VP2, VP3, VP4 再经过细心地重排,于是,产生出完整的病毒粒子。

小结一下,脊髓灰质病毒使用一个单个的 mRNA 分子,由这个 mRNA 翻译出的单个多肽链再经切割形成一些各自独立的病毒蛋白质,这与泡状口炎病毒不同,后者从(-)链病毒 RNA 合成一系列单顺反子的 mRNA 分子。

下面,我们看看披膜病毒,它像脊髓灰质炎病毒一样有(+)链病毒 mRNA,但它的生活周期却兼有脊髓灰质炎病毒和泡状口炎病毒两者的特征。

### 披膜病毒

披膜病毒 (togavirus) 含有一个单个的 (+) 链 RNA, 它编码 8 个蛋白质; 这个 RNA 分子包含两个 AUG 起始信号(见图 21-22)。病毒 RNA 不被加工,而用作 mRNA 翻译出一个多聚蛋白 I。合成起始于最靠近 5' 端的起始密码,接着,此

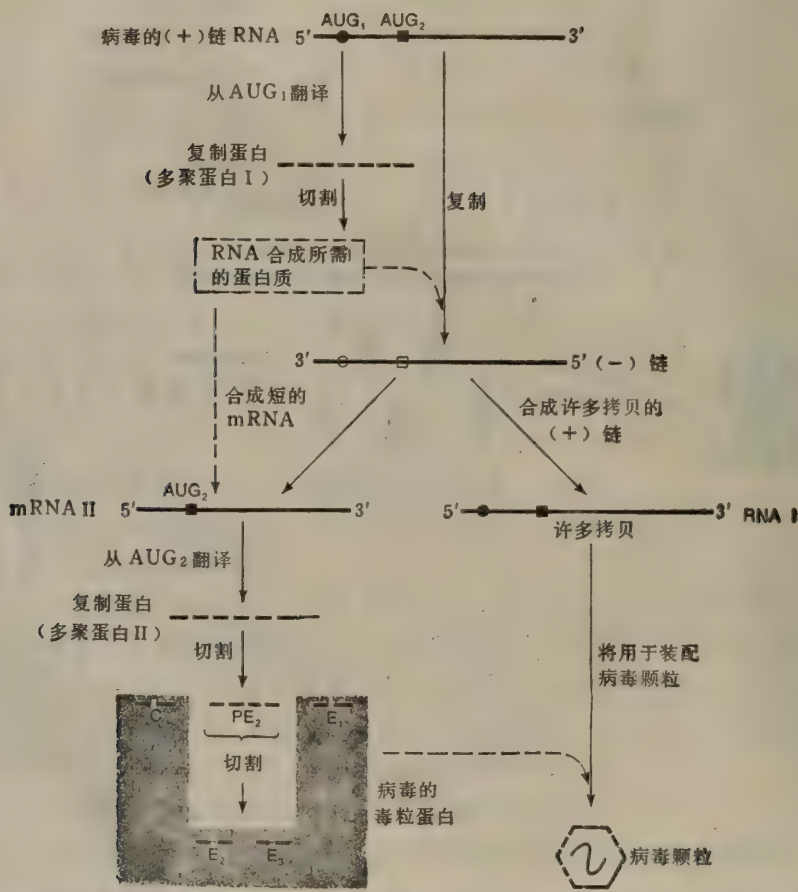


图 21-22 披膜病毒的生活周期。病毒链于早期被翻译产生复制蛋白。病毒链的(-)链复制于晚期进行拷贝,产生短的 mRNA 分子(缺 AUG<sub>1</sub>), 后者再被翻译成病毒粒子蛋白。



多聚蛋白质被切下,以产生出 RNA 复制所需要的蛋白质。此后这些复制蛋白质被用来从病毒 RNA 模板上合成出一个(-)链。这个(-)链起两个作用:(1)它是合成更多的、最终装配子代病毒粒的(+)链(RNA I,见图 21-22 中部)的模板。(2)它是形成一个包含 AUG<sub>2</sub> 起始密码而不是 AUG<sub>1</sub> 密码的 mRNA 分子(mRNA II)的模板。人们还认为从多聚蛋白 I 切下的若干蛋白质中的一种与 mRNA II 的产生有关。目前,还不知道 mRNA II 是从 RNA I 完整(+)链上切下来的,还是从一个内部的特殊的复制酶识别位点开始合成的。mRNA II 被翻译先产生多聚蛋白 II,这个大蛋白质再经切割产生蛋白质 C、PE<sub>2</sub> 和 E<sub>1</sub>。PE<sub>2</sub> 蛋白又进一步被切割产生 E<sub>2</sub> 和 E<sub>3</sub>。然后,蛋白 C、E<sub>1</sub>、E<sub>2</sub> 和 E<sub>3</sub> 一起装配形成外壳,其内装入 RNA。注意,两类暂时性的 mRNA (即“早期”病毒 mRNA 和“晚期”mRNA II)可以使披膜病毒调节合成的复制酶的量以及病毒颗粒蛋白质的量。也就是说,最初的病毒 RNA 的一个单拷贝被用来产生所需浓度甚低的催化性蛋白质,而 mRNA II 翻译产生的多聚蛋白 II 的合成则推迟,直到需要结构蛋白时才进行合成。由于有许多 mRNA II 分子,因此可大量制造结构蛋白。推测披膜病毒的产生比细小 RNA 病毒(脊髓灰质炎病毒)的产生有效得多。目前还没有观察到比这更高效的病毒产生过程,而且,披膜病毒与细小 RNA 病毒(只在哺乳动物细胞中生长)相比,有更为广泛的宿主(前者在昆虫和脊椎动物中都能生长)。事实上,经过一般的观察已看出,凡是复制和转录分开作用(就像在泡状口炎病毒中那样),对种种蛋白质合成的相对量以及合成时间的调控有加强作用的动物病毒来说,往往都伴随有宿主范围的扩大。

## 呼肠病毒

呼肠病毒(reovirus)(图 21-23)具有 10 个双链 RNA 分子,它代表着第三种类型的病毒遗传体系,它比迄今为止见到的其他病毒都要复杂。

呼肠病毒与其他病毒最大的不同,在于病毒颗粒的 RNA 分子在感染过程的任何时刻都不以游离形式从病毒颗粒中释放出来。

各自能编码一个多肽链的 10 个病毒 RNA 片段互相结合在一起,构成包裹于 20 面体外壳内的复合核蛋白核心的一部分(核酸部分)。当病毒进入细胞后便除去外壳,同时,构成核心组分之一的 RNA 复制酶被激活(图 21-24)。复制酶以 10 个 RNA 双链片段为模板,催化合成(+)链分子(即 mRNA)。核心内其他的酶催化每一个(+)链的戴帽过程。但是没有加上多聚 A 尾。迄今,呼肠病毒 RNA 是唯一已知的不具有多聚 A 尾的病毒 RNA 分子。新合成的(+)链从核心中伸展出来(图 21-25)。但是双链部分仍保持在核心内。每一条(+)链分别地翻译出一条单个的多肽链。这 10 个不同多肽链中的某几条经过切割形成最终的蛋白质。

10 个双链片段的产生是复制所必需的,这个过程是通过拷贝子代(+)链,而不是通过半保留地复制双链 RNA 完成的。这 10 个片段不多不少地包装在外壳中,形成新的病毒颗粒。在呼肠病毒中,组合过程是在调节下而且是在复制之前进行的。通过一种现在尚未知晓的途径,10 个(+)链中的每一个都与核心蛋白质结合在一起,形成一个核心前体(precore)。10 个(+)链以特定顺序依次被拷贝,这个过程可能是由核心蛋白质决定

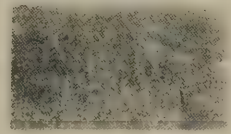


图 21-23 呼肠病毒的电镜照片(承 Robley Williams 提供)。

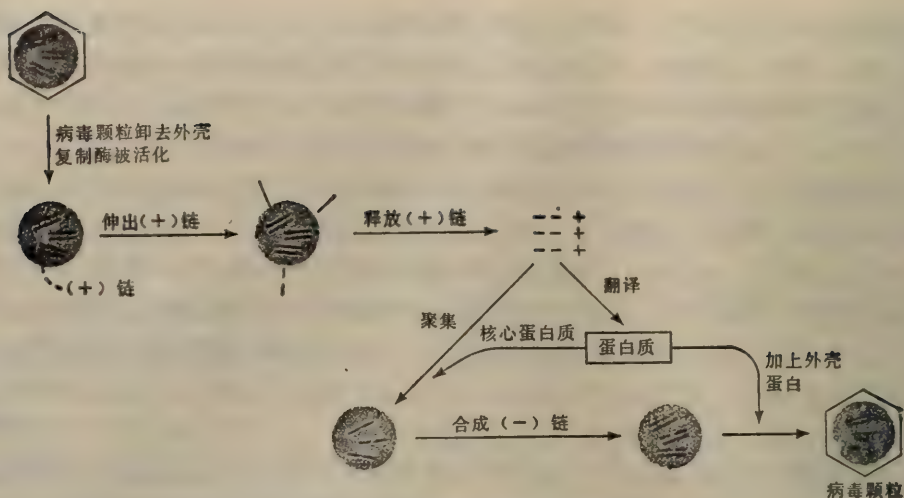


图 21-24 呼肠病毒的生活周期。

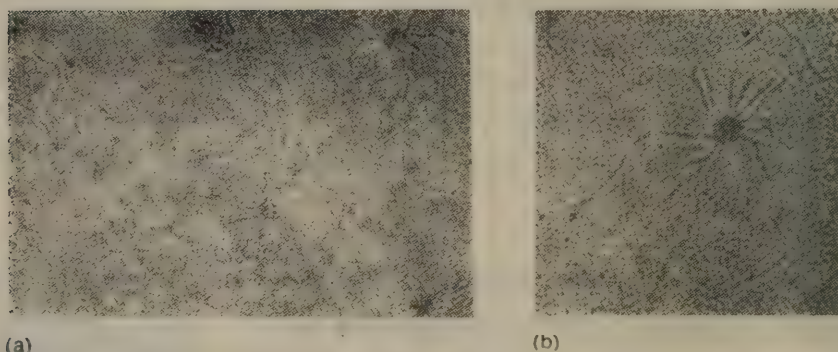


图 21-25 (a) 呼肠病毒的核心。(b) 将 (a) 的一部分放大。mRNA 分子从黑色核心中呈线状伸展出来 (N, M. Bartlett, S. G. Gillies, S. Bullivant 及 A. R. Bellamy 于 1974 *J. Virol.*, 14: 324)。

的。复制的结果形成了一个由一整套双链片段和核心蛋白质所构成的聚集体(注意,任何时候都没有出现游离的(-)链),然后迅速形成最终的核心。接着 20 面体外壳围绕着核心包装起来。

呼肠病毒生活周期的一个显著特点,是病毒颗粒的 RNA 的(+)链从未被征用过。其(-)链是合成所有子代(+)链的模板,子代(+)链则是合成双链病毒 RNA 的模板。

## 逆转录病毒

逆转录病毒 (retrovirus) 含有维系在一个复合体中的完全相同的两条主要的、大的及两条小的单链的 RNA 分子。逆转录病毒在所有已知病毒中是独特的,唯有它在生活周期中有一个逆转一般信息流的阶段 (即  $\text{RNA} \rightarrow \text{DNA}$ )。而不是像通常的机理那样 ( $\text{DNA} \rightarrow \text{RNA}$ )。在所有逆转录病毒的生活周期中,还有另一个不寻常的特征:要经过一个形成 DNA 中间体的必要步骤,此中间体就像噬菌体 Mu 那样插入到宿主的染色体中。大部分逆转录病毒都引起肿瘤,因而引起人们特殊的兴趣。在这一部分我们不讨论它们的致癌能力,只是考虑产生后代病毒的复制方式。



到目前为止,了解得最多的逆转录病毒是侵染鸡的劳氏肉瘤病毒 (Rous sarcoma virus, 简称为 RSV)。1964 年 Howard Temin 观察到, RSV 的增殖可以被氨甲蝶呤、阿糖胞苷 (DNA 合成的抑制剂)以及放线菌素 D(依赖于 DNA 的 RNA 合成的抑制剂)所阻断。在这一观察的基础上,他提出,这个病毒的生命周期中包括着一个 DNA 中间物 (DNA intermediate),他命名这个中间物为原病毒 (provirus)。这一假说提出了一个崭新的观点,即 DNA 可以依照 RNA 而合成出来。1970 年 Temin 和 David Baltimore 进一步分别证实并指出:在一些鸟类及鼠类的逆转录病毒中存在逆转录酶,这个酶能以单链 RNA 为模板合成出双链 DNA。从此以后,人们才改变了一直对此观点抱有的怀疑态度。

逆转录酶

逆转录酶具有三种酶活性:

- 1. 拷贝一个 RNA 分子,得到一个双链 RNA-DNA 杂合体,使用引物,以 3'-5' 磷酸二酯键的方式与 5'-脱氧核糖核苷三磷酸连接。
- 2. 可拷贝一个带有引物的单链 DNA,形成双链 DNA。
- 3. 降解 DNA-RNA 杂合体中的 RNA (有核糖核酸酶 H 活性)。

在体外,如果提供合适的引物,单链 RNA 可以逆转录出一互补的 DNA 链;接着,这一单链可以成为双链 DNA,如图 21-26 所示。

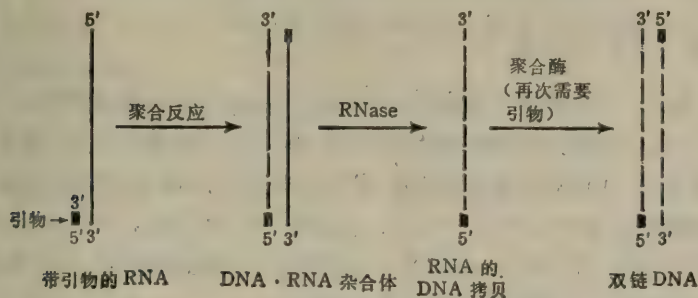


图 21-26 逆转录酶将单链 RNA 转变成双链 DNA 的假设过程。其中引物可能是 DNA,也可能是 RNA,它杂交到模板的 3' 末端。目前并不了解在自然界是否在模板的末端发生引物化过程。因此,真实的一系列过程必然比上述图示的要复杂得多。

如图中简单所示,需要一寡聚脱氧核糖核苷酸做引物,以避免线性 DNA 分子末端复制的问题。这一问题已在第十五章结合噬菌体 T7 的复制进行讨论。这种末端引物在自然界是得不到的,因此,需要一复杂得多的方案来加以解释,不久就会讨论 RSV 的方案。

劳氏肉瘤病毒的生活周期: 概貌

逆转录病毒的基本生活周期如图 21-27 所示,分为两个阶段。在第一阶段,病毒颗粒中含有的逆转录酶催化从单链病毒 RNA 到双链 DNA 原病毒的合成反应。在病毒发育过程中一个必不可少的步骤是双链 DNA 原病毒整合到宿主染色体中。人们设想原病毒以环状分子整合,但证据不足。插入 DNA 的基因顺序与病毒 DNA 中的次序完全相同,这一点和  $\lambda$  噬菌体不同,所以病毒 DNA 环化及整合位点必然是相同的。在第二阶段转

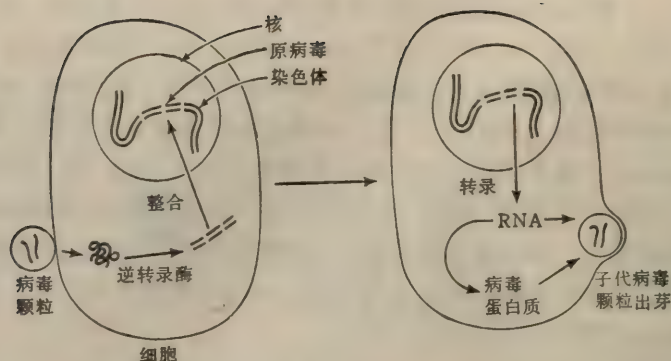


图 21-27 逆转录病毒的生活周期。病毒 RNA 进入细胞,通过逆转录酶转变为双链 DNA,双链 DNA 再整合到宿主染色体中成为原病毒。然后再转录出 RNA 及产生进行蛋白质合成所需要的 RNA。进一步再包装,产生子代病毒颗粒。

录发生了。转录子经加工修饰后成为 mRNA 分子,然后翻译成两个最终蛋白质和两个多聚蛋白质。这两个多聚蛋白质经切割得到 6 个以上的多肽链,其中两个蛋白质聚合成形成逆转录酶,四个蛋白质形成外壳 (capsid)。另外两个蛋白质形成被膜 (envelope)。通过一种尚不为人所知的方式,合成病毒 RNA。装配完成的病毒粒子以出芽方式,不断地从被感染细胞中凸出而脱落,未给宿主细胞造成明显的伤害。

以下部分将对劳氏肉瘤病毒的各种生物学特性作进一步介绍。

### RSV 病毒的 RNA 的性质

前面说过,RSV 不是仅含有单条的 RNA,而是有四条以氢键结合在一起的 RNA (图 21-28)。较大的一对(称为 38S 链或单体链)携带所有的遗传信息,而且完全相同。两条较小的链是宿主的 tRNA<sup>trp</sup> 分子。这种 tRNA 存在于所有鸟类的逆转录病毒粒子中。鼠逆转录病毒则携带宿主的 tRNA<sup>pro</sup> 分子。在后面我们将看到, tRNA 分子的位置对于复制是至关重要的。每一个 tRNA 分子的 3' 末端部分,通过氢键与 38S 链的距其 5' 端 101 个碱基对的被称为 pbs 的区域联系在一起。

劳氏肉瘤病毒 RNA 的另一个重要特征,是每个 38S 分子中含有两个相同的或非常相似的 21 个碱基的非编码序列。这些序列标记为 r (图 21-28)。r 紧紧地位于 5' 端帽子和 3' 端多聚(A)<sub>200</sub> 尾序列之邻。另外,还有两个非编码末端序列,即马上就要讨论的 u5

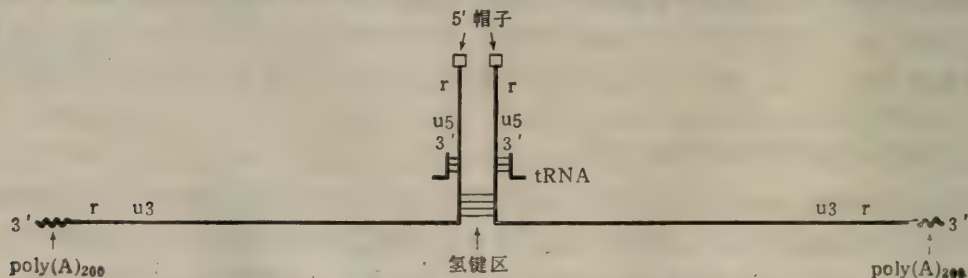


图 21-28 劳氏肉瘤病毒 RNA 的结构。注意,两个单体之间的氢键区形成两条平行的链,而不是形成两条反平行链。目前对这个非寻常区域了解甚少。r、u3、u5 代表一定的碱基序列,详见正文。这些序列的 DNA 拷贝用 R、u3 及 u5 表示。



和 u3。u5 和 5' 端 r 序列相邻, u3 和 3' 端 r 序列相邻。

### 病毒 RNA 向双链 DNA 中间体的转化

病毒 RNA 向双链 DNA 分子转化过程中,所有的中间物都只含有一条 38S RNA 链。推测这些中间物在病毒进入细胞质不久后便与 38S 部分分离。

RSV RNA 的合成早在转录和翻译之前就已进行。这是因为 RSV 本身携带着它自身的复制酶,即逆转录酶。如同许多复制酶一样,此酶也需要引物;逆转录病毒因为自身具有 tRNA 分子作为引物而解决了这个引发问题。引物位于 38S RNA 的 5' 端。结果,在生长着的 DNA 链到达模板 RNA 分子的 5' 端之前,只有一小段 DNA 能从 tRNA 的 3' 羟基聚合生成。

病毒 RNA 向双链 DNA 中间物转化的确切机理还不清楚,这里要说明的主要特性是病毒 RNA 碱基和 DNA 拷贝碱基之间的差异,如图 21-29 所示。人们发现, RNA 分子的 5' 端 r-u5 片段与 3' 端 u3-r 片段结合在一起所形成的 DNA 序列。U3-R-U5 出现在双链 DNA 分子的两端,即这个 DNA 分子是末端过剩的 (terminally redundant)。U3-R-U5 通常又写成 LTR, 即长末端重复。人们提出的反映转化反应的主要特征见图 21-30。它包括以下步骤:

1. 逆转录酶使一个 tRNA 分子从其 3' 羟基端向模板 5' 端延伸,生成一个 R'(U5) 的 DNA 拷贝。

2. 对应于 R'(U5)' 序列的 RNA 碱基被移去(可能是逆转录酶的另一种活性),从而使新合成的 DNA 的 R'(U5)' 可以提供待互补序列,成为形成氢键的一方。假设此时帽子和多聚 A 尾也被除去。

3. tRNA 与其互补序列之间的氢键分开,新合成的 DNA 分子与 RNA 分开,然后此 DNA 片段与 RNA 的另一端通过 r 和 R 之间的氢键连接在一起。

4. DNA 聚合酶在此 DNA 的 3' 端逐个地添加核苷酸,合成一条几乎完全的双链分子。

5. RNA 3' 端的 u3-r 序列被水解除去。

6. DNA 聚合酶在 RNA 的 3' 端逐个地添加核苷酸,完成双链分子,并形成第一个 LTR。

7. 所有 RNA 都被水解除去,此水解反应大概是由逆转录酶的核糖核酸酶活性催化的。

8. 第一个 LTR 解链,在分子的另一端形成 (pbs) (PBS)' 双链。这种配对为形成

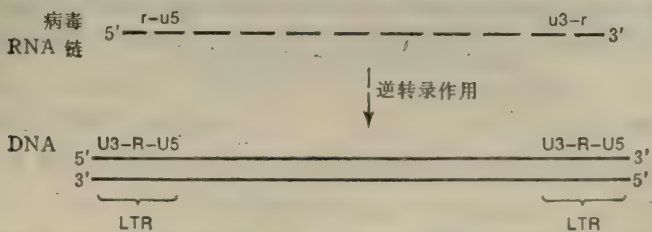


图 21-29 在一个 38S 的病毒 RNA 链上及其 DNA 拷贝上(用大写字母表示)的末端序列。DNA 中的 U3-R-U5 序列亦被称为 LTR 或长末端重复序列。

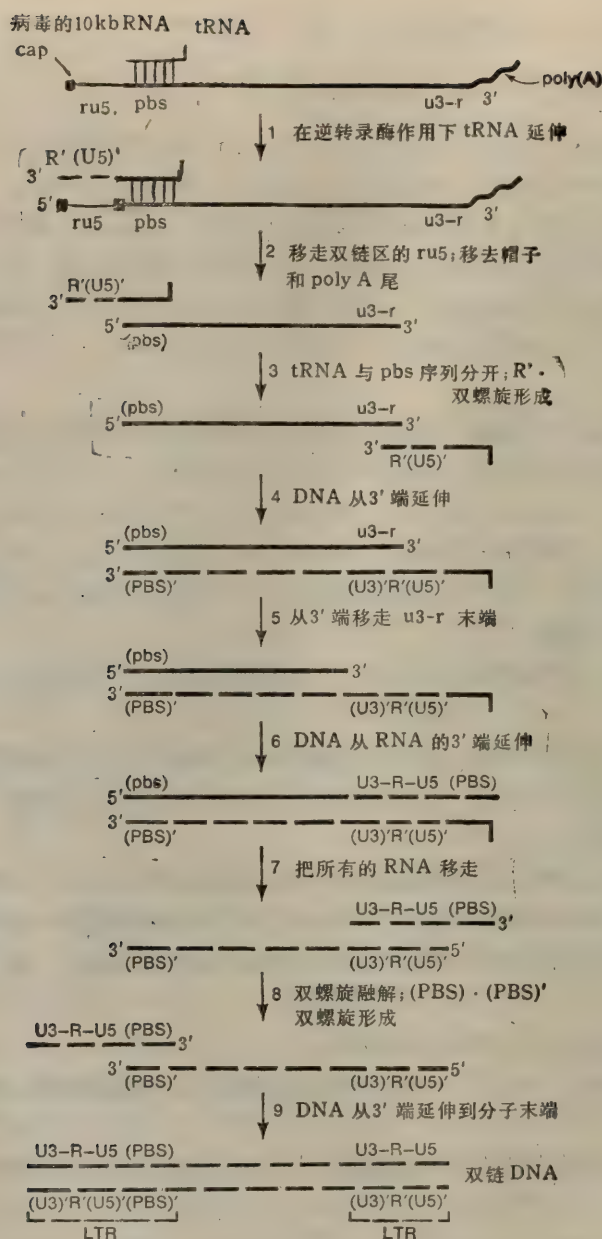


图 21-30 从劳氏肉瘤病毒 RNA 产生一个具有两个末端 LTR 序列的双链 DNA 分子的假设图解。pbs 序列与 tRNA 互补; 横线上面的字母与横线下面的字母以及数字分别表示 RNA 的序列与 DNA 的序列。“'”表示互补。RNA 用实线表示, DNA 用虚线表示。

第二个 LTR 提供了模板。

9. DNA 聚合酶向两个 3' 端分别添加核苷酸, 生成一个以 LTR 为两端序列的完全的双链 DNA 分子。

注意, 第 8 步生成的分子具有两个互补 LTR 序列, 因此第 9 步也可以是另一种方式, 即两个互补的 LTR 序列可以通过氢键形成双链 LTR。DNA 聚合酶催化一条链的



3' 端延伸到一定程度, 就可能合成出一个只带有单个 LTR 拷贝的双链的环形分子。这样的分子已经在被感染的细胞中观察到。不过线形分子的数目远远超过环形分子的数目。

图 21-30 中所示的最终分子以后整合到染色体中, 并在那里被转录。在生活周期中的某些时刻, 模板必须被拷贝以再产生病毒 RNA。显然这个过程应当发生在转录步骤, 但具体细节还不了解。

因为只有百分之一的 RSV 颗粒具有侵染力, 因此, 对于所观察的病毒, 难于将它在细胞内的种种结构与其生活周期联系起来。实验中必须使用比 100 个病毒粒子去感染一个细胞(100:1)更大些的比例去感染宿主细胞, 才能得到足够的感染细胞来进行研究。由于 99% 的病毒颗粒是非感染性(缺陷型)的, 所以, 如果每一个 RNA 分子在被感染细胞中都发生了某些改变, 那么所观察到的 99% 的分子完全有可能是从这些缺陷型颗粒发育而产生的, 这些缺陷型颗粒还能感染细胞。因此, 很可能不常见到的环形分子恰恰成为仅有的繁殖中间体。很清楚, 为了弄明白这种病毒的复制机理尚需进行大量的深入研究。

### 病毒 DNA 的整合

原病毒形成机理的任何细节都不清楚, 但是原病毒的结构却给人们提示了一种可能的而且有些令人吃惊的机理。用重组 DNA 技术把从被感染细胞中得到的原病毒克隆到大肠杆菌  $\lambda$  噬菌体上, 由此对整合进行研究。科学家分离了原病毒 DNA, 用限制性内切酶对其进行切割, 并测定宿主和病毒 DNA 的两个接点处附近的碱基序列, 然后把这些序列与未整合的病毒 RNA 端部的序列进行比较。这一实验没有采用 RSV, 而是改用 7 个其他的逆转录病毒作为实验体系, 用脾坏死病毒 (SNV, spleen necrosis virus) 的实

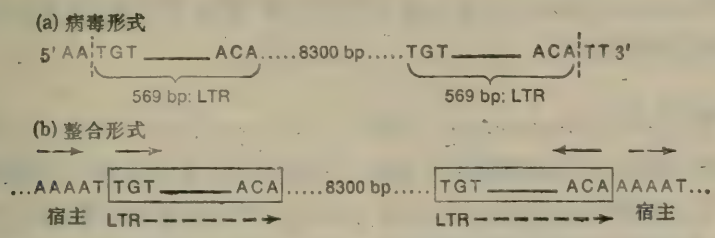
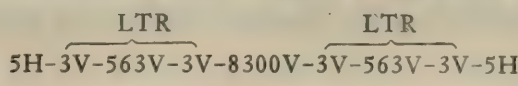


图 21-31 未整合的和整合的脾坏死病毒 DNA 序列。病毒 DNA 序列用实线表示, 宿主 DNA 用虚线表示。同种的箭头对 (实线、虚线及长虚线的) 代表相应的序列。如果箭头的指向相同, 那么它所代表的序列是直接重复的; 如果箭头的指向相反, 那么它所代表的序列是倒转重复。

验结果图示在图 21-31 中。(a) 行图示了无 5' 帽及无 3' 多聚 A 尾的病毒 RNA。之所以略去了这些末端区域是因为在逆转录复制中它们不被拷贝。可以看到, 长度为 569 个碱基的片段 (LTR) 位于每条链的末端, 而且每个 LTR 末端的最终三个碱基呈倒转重复。图中两条垂直虚线标明了原病毒的碱基界限; 也就是说, 末端的 TT 和 AA 不被整合。

(b) 展示了整合态的原病毒 DNA 链的结构。它有三点重要特征:

1. 569 个碱基长的序列是重复的。
2. 包含有 569 碱基重复序列的完整插入区域以 5'-TG 和 CA-3' 为末端。
3. 接点处宿主 DNA 是同向重复的 5 个碱基。因此, 整合的 DNA 序列可以写成



H 和 V 分别表示宿主和病毒。从上表示式中可以看到: (1) 病毒 DNA 两端是顺向重复的两个长序列 (LTR), 每个 LTR 的两端是反向重复的短序列 (3V)。(2) 病毒 DNA 插入序列两端为宿主 DNA 的顺向重复的短序列。在检查过的每一种属逆转录病毒中都发现这一格局; 唯一变化的是不同种病毒的短重复序列的长度和碱基序列有所不同; 并且, 至今观察过的所有整合的逆转录病毒 DNA 都以 5'-TG 和 CA-3' 为末端。这向人们提示, 可能存在一种适合于所有逆转录病毒的共同的整合机理。而当检查同一种属的、相互独立的一些原病毒时, 发现在每一个样品中, 同向重复的宿主短序列都不相同, 所以, 整合可能发生在宿主 DNA 的许多位点上。

原病毒及邻近区域的序列使人联想起转座子的性质。在第十九章中曾指出, 经过转座作用后, 可移动的遗传因子总是在其两个末端倒转重复序列 (2—1500 个碱基对) 的两侧再各接一个小的、同向重复的细胞 DNA (4—12 个碱基对), 而且, 酵母转座因子 Ty1、果蝇转座因子 Copia 以及噬菌体 Mu 的整合 DNA 也都以 TG 和 CA 为末端。难道它们的相似是偶然的? 太不可能。因此, 人们设想, 逆转录病毒的整合机理是类似于转座作用机理。实际上还没有证据表明逆转病毒的整合中确实发生了转座, 因为转座作用的标志是既要有靶序列的复制, 还要有转座子自身的复制。在逆转录病毒中, 染色体靶序列复制了, 但无人知道究竟是逆转录酶合成出的双链 DNA 分子本身整合进染色体, 还是这个 DNA 分子的拷贝从细胞质中转座到染色体的位点上。然而, 在序列分析数据的基础上, Howard Temin 对此提出看法: 逆转录病毒是从古代的可转座因子经衍生而形成, 并且有了复制和被壳化的能力。

一旦整合后, 原病毒 DNA 就待在宿主细胞染色体上而不被切除。原病毒的存在并不抑制细胞分裂, 所以子细胞也继承了整合的原病毒, 这些子细胞(以及后续世代的细胞)继续产生有活性的病毒颗粒。

### RSV 的转录作用

转录开始了 RSV 生活周期的第二阶段, 如图 21-27 所示。单一的转录子(标为 38S RNA)包含了所有的病毒基因(图 21-32), 一部分转录子保持其完整性以作为病毒 RNA 的来源以及作为合成多聚蛋白的 mRNA 分子, 多聚蛋白经切割生成 Gag 蛋白及 Pol 蛋白。许多 38S 分子被切割, 形成编码 env 和 src 基因的蛋白产物的 mRNA 分子。Gag 蛋白被切成病毒核心中的 4 个蛋白质。作为逆转录酶  $\beta$  亚基的 Pol 蛋白质, 也被切割成两个片段, 其中一个片段称为  $\alpha$  链, 它与  $\beta$  亚基结合形成有活性的逆转录酶。env 基因的蛋白产物是病毒粒子被膜中的病毒蛋白。Src 蛋白是致癌反应所必需的, 这一点将在癌病毒一节中讨论。src 基因突变后仍能产生有活性的病毒, 但是不能诱导向癌细胞的转化。

前面提到过, 病毒 DNA 的插入(实际上是逆转录病毒的活性)是病毒成功地完成其生活周期必不可少的。通过分析 LTR 及 38S mRNA 分子的碱基序列, 这种需求的原因之一便清楚了, gag 基因的编码序列开始于 U5 序列(见图 21-29) 右侧, 然而 38S mRNA 的启动子却在 U3 区域。所以, 在病毒 RNA 中, 启动子与编码区并不相连。当 U3 与 R-U5 序列相连形成 LTR, 就把启动子带到编码序列的上游位置, 于是 38S mRNA 得以合成。



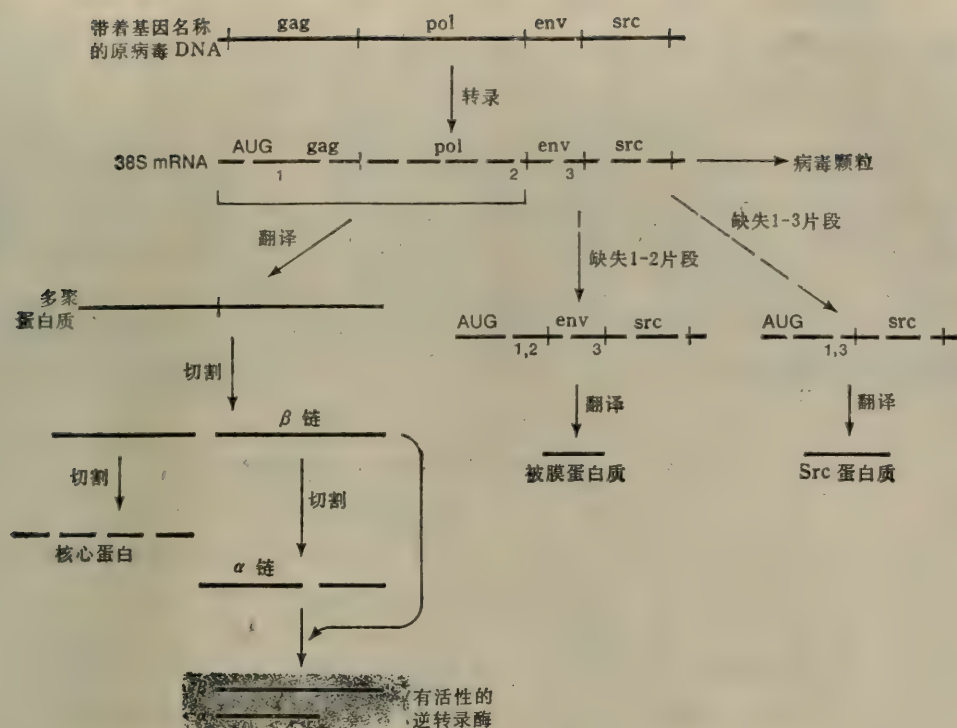


图 21-32 劳氏肉瘤病毒的基因顺序。并图示了其 mRNA 分子(虚线)及其蛋白质。

## 动物 DNA 病毒

有人可能认为 DNA 病毒会比 RNA 病毒少一些麻烦,因为 DNA 病毒的复制及转录应遵循 DNA 噬菌体的方式,而且 DNA 病毒不需要逆转录酶那样的特殊酶。理论上讲,一个 DNA 病毒只含有一个 DNA 分子,此 DNA 编码一或两种外壳蛋白。微小病毒(parvoviruse)确实遵循这种简单的方式。但是我们将看到大多数 DNA 病毒的生活周期要复杂得多。有些病毒进化形成 100 多个基因,比起只有区区几个基因的病毒来说具有较大的多变性,从而成为极端复杂的结构。本部分我们仅仅介绍多种 DNA 病毒中的很少几种,即挑选那些生活周期较为明了的几类 DNA 病毒进行讨论。这些病毒包括微小病毒、乳多孔病毒(papovaviruses)及腺病毒(adenovirus)。为了讨论已经充分研究过的疱疹病毒(herpesviruses)和痘病毒(poxviruses),例如牛痘苗病毒(vaccinia)和及乳头病毒(papilloma),读者应参考在本章后所列参考文献。

### 微小病毒

微小病毒有两类:自主型的和缺陷型的。自主型的病毒凭借它们自己及宿主的基因产物便可产生后代。而缺陷型的病毒还需要一些外加的蛋白质,它们既非病毒本身的核酸又非宿主的核酸所形成的基因产物。这类病毒只有在宿主被另一种含 DNA 的病毒,例如腺病毒感染时才能复制。推测腺病毒可能对它提供了必需的蛋白质。

微小病毒含有一个非常小的单链 DNA 片段。这个 DNA 一般含有 4 800 个碱基，它编码三个蛋白质，分子量分别为 61 000，64 000 和 83 000。这三个分子量的总和比 4 800 个碱基的编码容量（若在一个阅读框架中它所编码的蛋白质的最大分子量应为 176 000）超出了 32 000。因为它是使用了三个阅读框架（目前，除了大肠杆菌噬菌体  $\phi$ X 174 之外，微小病毒是唯一已知的具有重叠基因的病毒）。生成的这三种病毒蛋白只用于建造外壳。微小病毒利用宿主的酶去进行转录和复制。人们注意到这种现象，即这些病毒只是在宿主 DNA 复制时才能复制，而且病毒 DNA 的复制只在宿主细胞核中进行。推测，病毒复制所必需的一些宿主的酶只在彼时彼地才能得到。

不论自主型的还是缺陷型的微小病毒，虽然细节有所不同，其复制过程都很相似。了解得最多的是腺联病毒（adeno-associated virus, AAV）——一种缺陷型病毒的复制，这

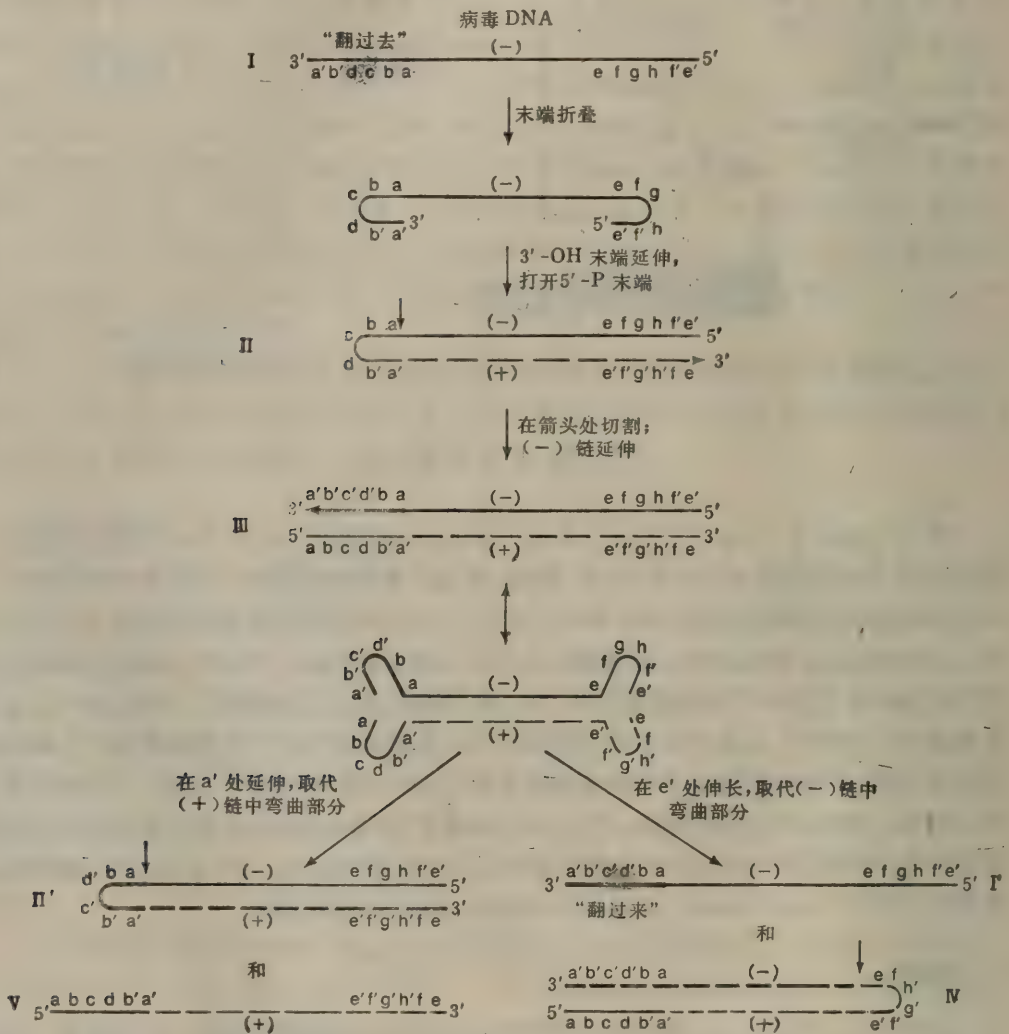


图 21-33 微小病毒 DNA 的复制图示。这个 DNA 在子代(-)链中产生一个倒转的序列(图中阴影部分)。短的垂直箭头指出发生专一性位点切割的位置,产生 3'-OH 基,从这个 3'-OH 基处可以再生长出一段新链。图中所示的各种分子可以进一步生长,延伸,并将能产生在末端片段中有倒转的其他序列。



部分所描述的图示正是该种病毒。

只有一个单链 DNA 的微小病毒同样面临着通常的问题——为 DNA 合成的起始怎样提供引物。然而，其 DNA 碱基序列的排列使其可以自行引发复制。这一单链含有两个末端反向的重复序列，其末端迴折，形成终端发夹环，如图 21-33 所示，其中一个发夹环具有 3'-OH 末端碱基对。这提出了引发的完美备件，图中展示了引发后 DNA 复制的各个阶段，这一复制模型有时称为发夹转移模型 (hairpin transfer model)。首先，3'-OH 端延伸，当伸长的链到达 5' 端时，发生位移反应，5' 端发夹环打开，从而保证链伸长到 5' 端。在这个阶段，分子是一个完整的双链发夹 (II)。然后在图中小箭头所示处断开，继续延伸，生成一个双链线性分子 (III)。接着，末端发生解链，各个末端的反向重复序列分别形成两个短的发夹环，又出现 3' 端发夹环，3'-OH 引物又可用于链的延伸。这样的末端有两个，一个或两个都可被征用。若两个同时征用，将出现结构很大的分子 (已经观察到)。在这里我们看看仅用一个活性引物的分子。从中间物左边 3' 端延伸得到分子 II' 和 V。从另一个 3' 端延伸得到分子 I' 和 IV。如同亲代分子一样，分子 I' 是负链，但其短反向重复片段即 3' 端序列为 a'b'c'd' ba(III)，而亲代中为 a'b'dcba(II)。倒位同样也存在于其他三个分子中，这两个序列称做“翻过去” (flip) 和“翻过来” (flop)。正是在后代病毒中发现有这两个序列存在，支持了图中所示的复制方式。复制过程可能在箭头后面对分子 II' 和 IV 切割，从而生成类型 III 分子——双链线性分子。它可能再转变成四重发夹中间体，分子 V 也同样依照途径 I-II-III 进行复制。然而，正像图示模型，起始分子 I 为一条(-)链，而分子 V 却为一条(+)链。在某种范围，它也被包装，而且可有这些分子的颗粒也是有感染性的。(+)链的复制与图示的方案一样，(+)(-)链均可生成。这样，在一组大量的病毒样品中，可以发现(+)链病毒及(-)链病毒两种，而且在“翻过去”和“翻过来”两种构型中都有。但在自主型的微小病毒中还未发现“翻过去”和“翻过来”颗粒。所以它们可能启用另一种复制途径。

微小病毒的转录方式目前还一无所知。

### 乳多孔病毒：多瘤病毒和 SV40

多瘤病毒和猿猴病毒 40(SV40) 是两个研究得较清楚的乳多孔病毒(图 21-34)。它们在感染某些宿主时都非常高效地致病；在另一些宿主中，它们是烈性的，以破裂被感染细胞和释放子代病毒为结局。在这一部分我们只讨论其裂解周期。

#### 乳多孔病毒颗粒及其 DNA 的性质

所有乳多孔病毒颗粒都含有一个环状 DNA 分子的裸露 20 面体，纯化后，DNA 呈超螺旋状，外壳包括一个主要的外壳蛋白 VP1 和两个较次要的外壳蛋白 VP2 和 VP3。除了组蛋白 H1 之外，此病毒的 DNA 能与所有宿主组蛋白结合，这些组蛋白引起 DNA 分子紧缩而形成真核染色质特有的珠状核小体结构(见图 21-35 和第六章)，这些珠状结构被称为小型染色体 (minichromosome)。由于组蛋白的结合引起双螺旋轻微解链，而且组蛋白在复制中自始至终与 DNA 结合，所以在一圈复制结束时，子代分子也处于轻微的解链状态。但在小型染色体中，由于 DNA 围绕组蛋白旋转而使其轻微解链状态变得不易察觉。然而，当 DNA 经纯化，在去蛋白一步除去组蛋白后(如经酚抽提)，DNA 就

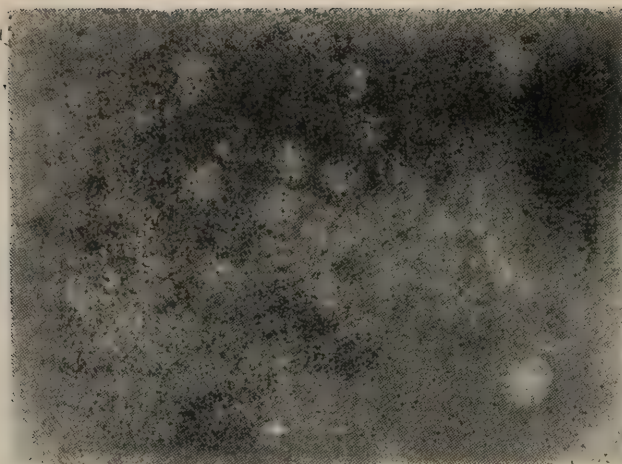


图 21-34 SV40 病毒的电镜照片。

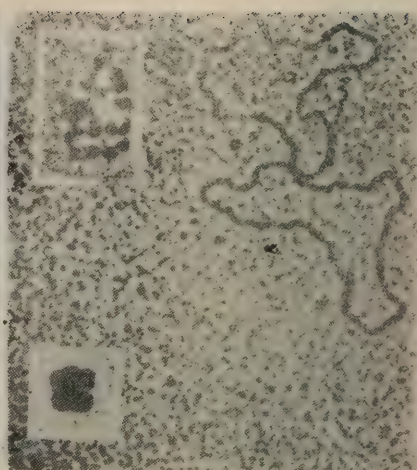


图 21-35 SV40 DNA 借助宿主的组蛋白而有不同程度浓缩。

呈现超螺旋构型。

多瘤病毒及 SV40 的 DNA 分子顺序已完全测出;分别都含有 5224 个碱基对。

在下面将讨论 SV40。讨论结束时将对 SV40 和多瘤病毒之间的微小差异做一些评述。

#### SV40 的生活周期

SV40 在猴肾细胞中以裂解方式生长。如同许多病毒侵染一样, SV40 的生活周期可分为早、晚两期。在其生活周期中合成出五种病毒编码的蛋白质(见表 21-6)。侵染后先出现一个长的(12 小时)潜伏期,此期间病毒 DNA 是不活化的,它们逐渐向宿主细胞核迁移。不论转录还是复制都无法在核外进行,大概所需的多聚酶只存在于细胞核中。

表 21-6 SV40 病毒编码的蛋白质

蛋 白 质	合成时间	功 能
T 抗原	早	起始 DNA 复制
t 抗原	早	未知
VP1	晚	主要的外壳蛋白
VP2	晚	次要的外壳蛋白
VP3	晚	次要的外壳蛋白

在感染的早期,主要合成的病毒蛋白质为 T 抗原和 t 抗原\* (图 21-36)。在早期 mRNA 合成后,宿主受到刺激合成一些酶——DNA 聚合酶、胸腺嘧啶激酶、dTMP 激酶和 dCMP 脱氨酶。不久会看到,这些宿主酶也都是 DNA 合成所必需的, T 及 t 抗原同样是病毒 DNA 合成所需要的。几个小时后,宿主 DNA 也受刺激开始合成,但不清楚这是病毒生活周期不可缺少的,还是不可避免的不同步现象。大约在病毒和宿主复制酶合成后 12 小时,开始合成病毒 DNA。不久,晚期 mRNA 合成了,显然病毒 DNA 的

\* T 抗原及 t 抗原的名称由免疫荧光显微镜观察而得出,这些蛋白质位于那些被病毒感染而转变为肿瘤细胞的细胞核之中。



大约 3 天之后,被侵染的细胞裂解,每个细胞释放出  $10^5$  个病毒颗粒,裂解很可能是由于很大的病毒粒子结晶体状排列引起的机械损伤所致,而不是由裂解酶引起的。

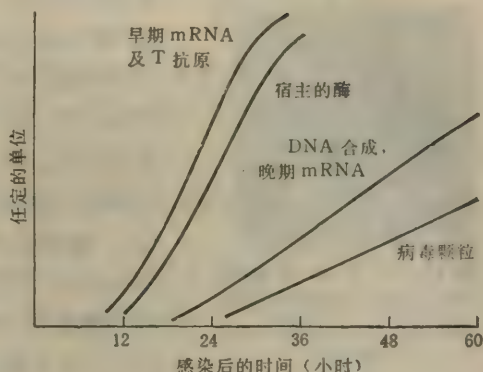


图 21-36 SV40 生活周期中发生各种过程的时刻。

## SV40 DNA 的物理图谱

图 21-37 展示了 SV40 的物理图谱, 其中指出了复制起点的位置和 mR

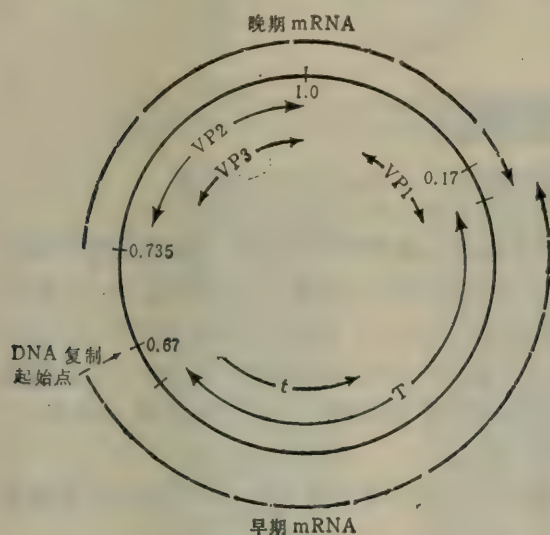


图 21-37 SV40 的物理图谱。图谱示出了 T, t, VP1, VP2, VP3 5 个基因及复制起始点的位置, 以及早期 mRNA 分子和晚期 mRNA 分子。

的结合位点(图 21-38)。T 抗原是一个必需蛋白质,假若没有 T 抗原,那么 SV40 的 DNA 合成就无法起始。这首先是由如下事实所阐明的,即温度敏感突变株 A(Ts) (T 抗原的基因被称为 A 基因)在高温下不能起始 DNA 合成。T 抗原还是一个 ATP 酶,它在起始过程中的确切功能还不知道,不过它的其他功能将在有关癌病毒部分进行探讨。

一个有趣的观察表明,复制起始点必定具备不寻常的结构,当单链DNA的S1核酸酶作用于超螺旋病毒DNA(已脱去蛋白质)时,在DNA复制起点处形成一个单一切点\*。此酶只作用于单链DNA,对于非超螺旋DNA则是无活性的。在第四章中曾指出过,超

这个图是经过下列几步建成的：用许多限制性内切酶切割病毒 DNA，把片段依次排好形成一个限制性内切酶图谱，再用片段与从被侵染的细胞中分离到的、标记了的 SV40 mRNA 杂交。又在体外翻译体系中使用纯化的 mRNA，鉴别出对应于上述 mRNA 分子的基因产物。并且用上法鉴定了特殊 mRNA 分子所合成的蛋白质。EcoRI 酶在 SV40 上只有一个切点，其位置被选为此图谱的参考点或零点。

## SV40 DNA 的复制

除了起始的机理外, SV40 DNA 复制已研究得相当清楚了, 复制起始点位于图谱 0.67 位置处。这是 T 抗原

\* S1 核酸酶是一种从真菌 *Aspergillus oryzae* 中分离得到的核糖内切酶。它作用于单链 DNA、RNA 及变性泡的单链分枝上。它是分子生物学中的一种很有用的试剂。

螺旋 DNA 分子中的很大一部分碱基对可以在任何时刻被解开,而且这些融解区又可以

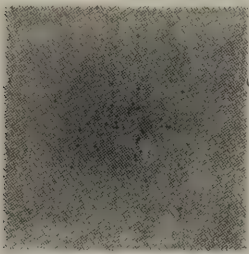


图 21-38 SV40 的抗原分子。用负反差电子显微镜观察(承 Robley Williams 提供)。

被连续地改变,因此也许可以设想,如果 DNA 是超螺旋的,在大部分时间里, A + T 富集区域保持单链形式。从 S1 作用的结果可以推测, SV40 复制起始区是 A + T 丰富的。当 SV40 DNA 碱基序列全部测出之后,这一想法得到证实。图 21-39 画出了复制起始区的碱基序列,这个区域的确具有一个不寻常的结构——含有 27 个碱基对的 G-C 富集序列。在这个序列中围绕一个 G-C 对,在它的两侧各有 13 个 G-C 碱基对的反向重复序列(即图中方框中的部分),其一边的外侧还有 17 个富集 A-T 碱基对序列。一个 T 抗原的四聚体首先结合在这个富集 G-C 的回

文结构上,然后,另外两个 T 抗原四聚体协同地结合到余下的 G-C 区和接着的两翼 A-T 区上。复制从作为两个反向复制序列的对称轴的 G-C 碱基对开始。有意思的是,这一位点与晚期 mRNA 合成起始位置也很近。不久将会讨论这个问题。

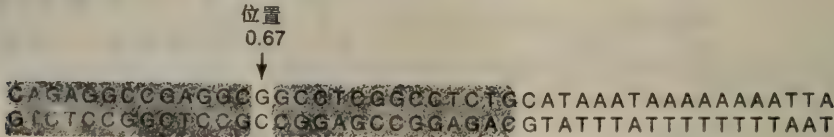


图 21-39 SV40 复制起始点的碱基序列。在 0.67 处的一个 G-C 对的两侧各有 13 个碱基对的倒转重复序列。

由 B. Hirt 建立起来的 DNA 分级分离技术,使得病毒 DNA 复制的研究成为可能。B. Hirt 提取法包括用去污剂十二烷基硫酸钠(SDS)和高摩尔浓度的 NaCl 依次地处理被感染的细胞,0℃恒温过夜后,含有 SDS、NaCl、细胞蛋白质及高分子量 DNA(此处是宿主 DNA)的复合体被沉淀下来,上清液中含有大部分的病毒 DNA 而几乎没有宿主 DNA。这些病毒 DNA 既有复制型的也有非复制型的,可以直接用电镜检查。人们观察到 SV40 DNA 复制的以下重要特性:

1. 复制是双向的,和大肠杆菌的  $\lambda$  噬菌体一样,两个复制叉在相对于起始的位点会合而停止复制。
2. 部分复制的 DNA 分子含有完整的亲代链,未复制区域仍然是超螺旋结构(见图 21-40)。在这一中间体内,杂合的子代环展开,形如蝴蝶,故有时称为蝴蝶分子,某些亲代双螺旋分子必须发生一些解链过程。如果没有解链作用去解开扭力矩,整个结构就会成为紧紧缠绕的一团。扭力矩的释解是由拓扑异构酶催化的,不过这一特殊酶还未被最后鉴定。
3. 在病毒生活周期的晚期,发现大量长的串联状的 SV40 DNA,这些分子可能是滚动环的分支,不过滚动环还几乎未被分离到\*。可以用这种串联体分子去侵染猴肾细胞;人们发现在宿主细胞中它们被切成单体,然后被用于产生子代病毒。形成串联体的重要性尚不清楚,但有可能像大肠杆菌的  $\lambda$  噬菌体那样,串联体是病毒 DNA 的前体。

\* 偶然见到过带有一小段线性分枝的环状分子。然而,如果这个分枝的长度小于环状分子的周长,那么就很难确定究竟是滚动环分子,还是复制叉处断裂的  $\theta$  型分子。



除了起始这一步,已在体外进行了 SV40 DNA 复制的研究(就是说,实验者必须从一些已部分复制的分子开始研究)。这些体外研究给出了以下事实:

1. 一条子链是连续合成,另一条则是不连续合成的。
  2. 延滞链的每一个前体片段用 10 核苷酸链为引物合成,此引物是由与大肠杆菌中的引物酶相像的酶所催化合成的。引物的碱基序列变化很大,说明并不存在识别某专一性序列的过程。
  3. 不像大肠杆菌那样,引物的除去过程与前体片的连接过程并不偶联,引物的除去是在早得多的时候进行的。
  4. 聚合酶是 DNA 聚合酶  $\alpha$ , 它是主要的宿主 DNA 复制酶。
- 目前,纯化的 T 抗原正在体外起始 DNA 的合成中试用。

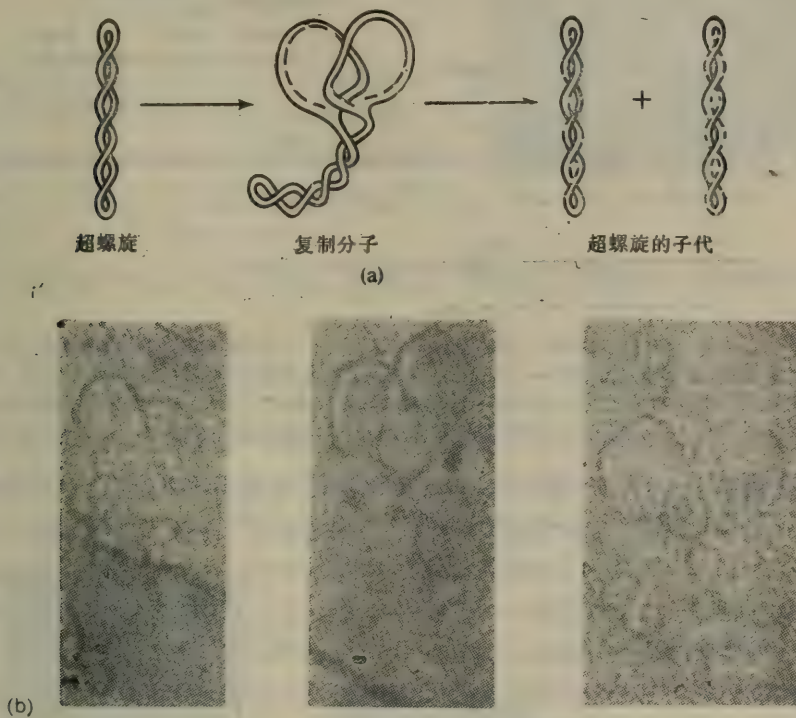


图 21-40 微小病毒 DNA 的复制。(a) 复制分子在其不复制的部分保存它的超螺旋形式。(b) 三个复制分子。

### SV40 感染过程中的转录

SV40 DNA 的转录是一个复杂的过程,涉及到时间的调控(早 $\rightarrow$ 晚)和广泛的 mRNA 的加工过程。在这些过程中,选择性地移去起始和终止密码。早、晚期 mRNA 分子分别从 DNA 分子的不同链上转录下来;它们合成的方向是相反的。而且,它们可能从毗邻于 DNA 复制起始点的几个相近的位点处起始转录(见图 21-37)。图 21-41(a)指出,编码早期 mRNA 分子是从图谱位置 0.66 左边的一个启动子开始合成的。它负责编码 T 抗原和 t 抗原。一个 AUG 密码子负责起始上述这两个蛋白质的合成。这条 mRNA 通过两种不同修饰途径产生两个单顺反子 mRNA 分子,一个编码 T 抗原,另一编码 t 抗原。

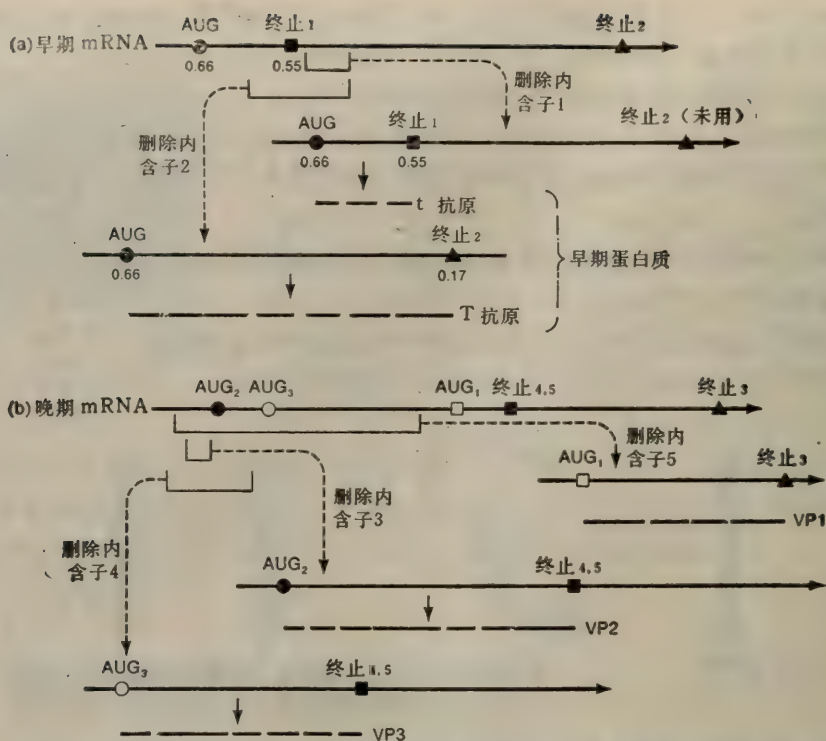


图 21-41 SV40 的早期 mRNA 及晚期 mRNA 的加工。—表示将要删除的内含子；从上往下的虚线箭头表示从未加工的 mRNA 分子转变到删除了内含子的、已加工的 mRNA 的过程。空心与实心的小方块、圆形及三角形表示在加工过程中各级 mRNA 分子上相同的起始密码与终止位点。mRNA 用实线表示，蛋白质用长虚线表示。

在第一种加工模式中，内含子 1 被切去，产生 mRNA，于是，翻译就从 AUG 开始到终止点 1 停止，产生出 t 抗原。而 T 抗原则是通过移去内含子 2 进一步翻译至终止点 2 而形成的。内含子 2 与内含子 1 的 3' 端相同，但其 5' 端 mRNA 的 5' 末端更近。切除这个大的内含子时移去了终止点 1，所以翻译一直进行到终止点 2，从而得到比 t 抗原大得多的 T 抗原。注意 t 抗原和 T 抗原的氨基末端的片段完全相同。

晚期 mRNA 编码三个蛋白质：VP1、VP2 和 VP3。剪接的方式则决定了哪一个蛋白质被翻译，使用重叠的阅读框架使病毒能在相当小的一段 DNA 中储存大量的信息。只有一个晚期 mRNA 从图中 0.735 位置起始转录而成。图 21-41(b) 示出由三种剪接方式所得到三种单顺反子 mRNA 分子，以及由此所合成的三个蛋白质。注意，其中没有一种蛋白质是从未经剪接的 mRNA 分子上翻译出来的。VP2 和 VP3 从同一个阅读框架翻译出来，两者具有相同的羧基末端，所以，VP3 实际上是 VP2 的一个片段。当加工酶作用除去编码 VP2 的起始密码子 AUG<sub>2</sub> 及其随后的一段 RNA 后，AUG<sub>2</sub> 后面的另一个起始密码子 AUG<sub>3</sub> 就起始翻译，产生 VP3。VP1 的阅读框架与 VP2 和 VP3 的不同，它是切去一大段 RNA（其中包含有 VP2 和 VP3 的起始密码子）后形成的。其中唯一剩下的起始信号 AUG<sub>1</sub> 就用于翻译 VP1。

注意 SV40 的这些基因产物都不是由多聚蛋白切割而得到的。

多瘤病毒 DNA 的复制方式与 SV40 DNA 的相同，转录也几乎相同，不同之处是多



瘤病毒具有三种(而不是两种)加工早期 mRNA 分子的方式,从而使它能合成第三个蛋白质,称为中等 t 抗原。这个蛋白质的功能将在涉及到瘤病毒部分再作叙述。

### 成熟病毒颗粒的存在部位

SV40 是致液泡病毒,也就是说,在侵染晚期,被侵染的细胞积累着一些称之为液泡(vacuoles)的袋状物,它们有外膜,并含有液体。病毒 DNA 在细胞核中合成之后,转移到合成蛋白质的细胞质内。病毒的组装发生在勾画出液泡轮廓的细胞质膜与细胞质膜之间的地方(图 21-42)。再晚些时候,当细胞解体时,膜破裂,病毒颗粒从细胞中释放出来。

## 大型 DNA 病毒

到目前为止,已介绍了一些只有少量基因产物,而且 mRNA 分子数目很少的一些病毒。还有一些较大的病毒,正如所预想的那样,它们具有极其复杂的生活周期。其复杂程度在任何情况下都不敢说已将它们的生活周期搞清楚了。

在这一节我们将对以下三种病毒:腺病毒、单纯疱疹病毒和牛痘病毒进行提纲挈领地讨论。

### 腺病毒

腺病毒是研究得较透彻的一大组裸露 20 面体病毒(图 21-43),每个病毒含有一个双链 DNA 分子,不同病毒的 DNA 的分子量范围为  $23-30 \times 10^6$ 。DNA 编码至少 20 种蛋白质,其中 15 种包含在病毒颗粒内。4 种病毒蛋白与 DNA 分子相联——第一个与磷酸

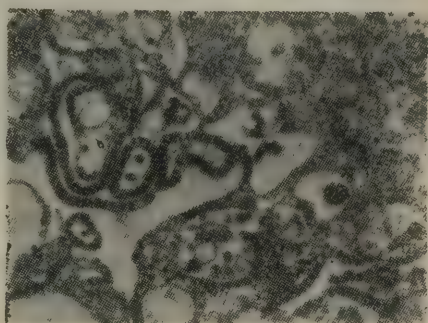
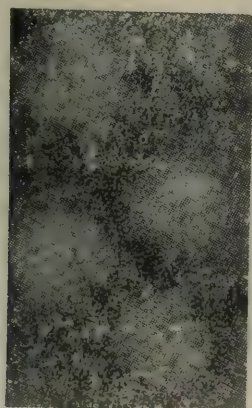
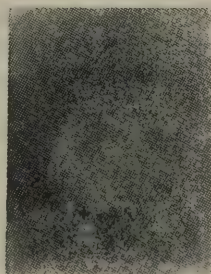


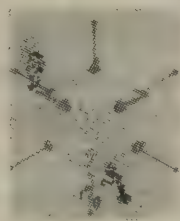
图 21-42 细胞膜中成熟的 SV40 颗粒。



(a)



(b)



(c)

图 21-43 腺病毒颗粒。(a)几个病毒颗粒的电镜照片。系用反差技术摄制。每一个病毒颗粒由 252 个壳体(在显微照片中每个壳体呈白色圆球状)排列形成一个 20 面体。240 个壳体是有 6 个相邻面的六体[见图 (c)中一个个白色球状体],12 个顶部壳体是具有 5 个相邻面的五体;每一个顶部壳体由一个碱基和一个突起组成,像一个长钉子。(b)一单个病毒颗粒的电镜照片,可以看出长钉状结构。(c)一个人工制作的病毒颗粒的塑料模型。其上白球表示六体,黑球表示五体。

基团作用;第二个使 DNA 形成多叶形,而且使 DNA 结合在被膜的 12 个亚基上;第三个共价地结合到 DNA 分子的 5' 末端;第四个的功能还不清楚。这些连接蛋白在后面还将详细地讨论。腺病毒在被侵染细胞中形成的大晶体排列于细胞核中(图 21-44)。

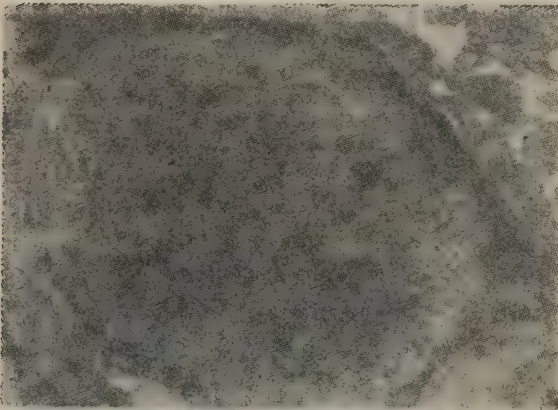


图 21-44 在受感染的 L 细胞的细胞核内的腺病毒结晶。

人们对于腺病毒的兴趣最初很大程度是由下列观察结果所激起的:引起人和猿猴呼吸道感染的腺病毒对于新生大鼠和小鼠是致癌的。而且,腺病毒还能使组织培养中的啮齿类动物的细胞成为转化细胞(见后面部分)。

#### 腺病毒 DNA 的末端

腺病毒 DNA 是一线性分子,它具有的反向重复末端一直精确地延伸到分子的尾端(图 21-45)。在腺病毒家族中这一重复序列的大小范围是 103—162 碱基对。不同病毒 DNA 末端的最前面 50 碱基对,具有相当大的同源性。这些序列中 A-T 极其丰富。由于腺病毒 DNA 的其他部分的序列变化多端,腺病毒上述序列中的明显的同源现象说明了这一部分的重要性。的确,它很可能被用于 DNA 复制的起始。

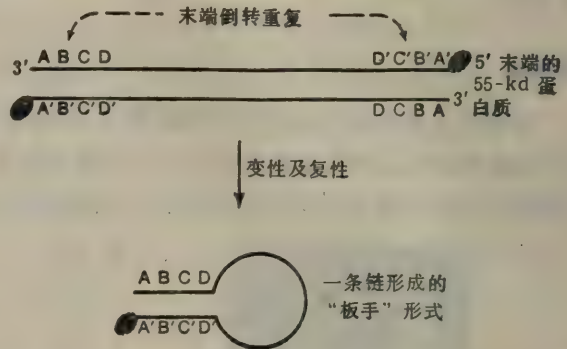


图 21-45 腺病毒 DNA 的结构及“扳手”状形式(panhandle form)。后者是在实验室中经变性又复性而形成的结构。

后来发现了腺病毒 DNA 末端的另一个重大性质,即当人们从病毒颗粒中分离 DNA 时, DNA 自发地形成环状分子和寡聚体分子。这使人想起了  $\lambda$  噬菌体 DNA: 它的互补单链末端可以形成这些结构。但是,腺病毒 DNA 并不具有单链末端,当使用蛋白酶处理腺病毒 DNA 后,它便失去形成这些结构的能力。对腺病毒 DNA 形成环状结构及寡聚体结构,解释为腺病毒 DNA 的每一个 5' 端共价结合着一个端基蛋白质(55kd 蛋白质),这些蛋白质可以形成二聚体甚至更高的多聚体(图 21-46)。蛋白质与 DNA 是通过 55kd 蛋白质中丝氨酸残基的 OH 基与每一个 DNA 链末端脱氧胞嘧啶的 5' 磷酸基之间的磷酸二酯键连接在一起的。这个蛋白质在 DNA 复制中起着重要作用[在其他生物体中也观察到端基蛋白质。研究得最多的是枯草杆菌噬菌体 29, 其他还有微小病毒 H1、小鼠小型病毒(minute virus)及乙型肝炎病毒]。



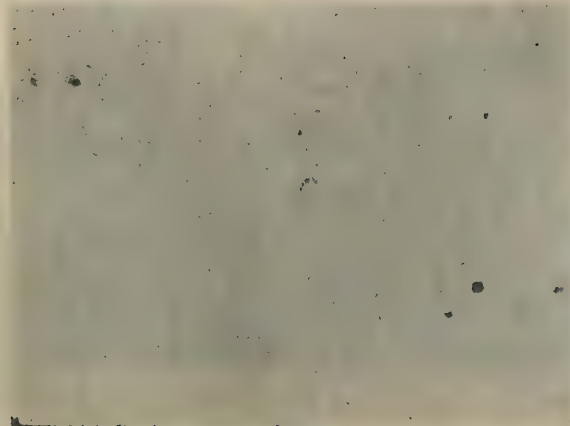


图 21-46 证明腺病毒的 55kd 蛋白质具有形成多体作用的电镜照片。图示各自带有两个端基蛋白质的三个腺病毒 DNA 分子。55kd 蛋白质先使两个 DNA 分子发生二聚作用,再通过蛋白质-蛋白质相互作用使三个 DNA 分子聚合到一起形成图中所示的三体分子。

### 腺病毒 DNA 的复制

腺病毒 DNA 复制的研究表明:

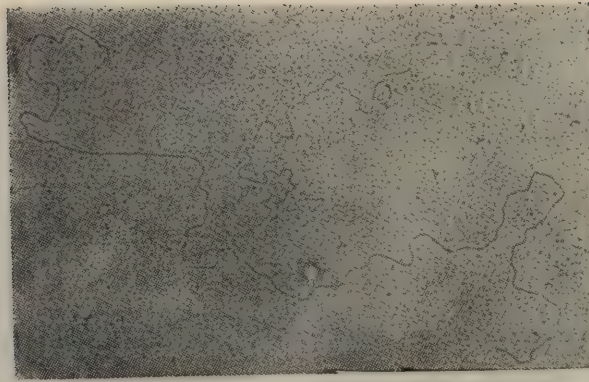
1. 复制的起始可以从 DNA 分子的任一末端开始。因为两个末端是反向重复,无法区别它们,所以起始以相同的频率发生在两端。而且,有些分子在两端同时起始复制。起始不需要 RNA 引物,这一点后面还要讨论。

2. 生长的子链延续到分子的另一端,在这个过程中,一条母链被一条新生的子链取代。在两端同时起始的分子中,两条母链都被取代,这意味着复制中间物具有伸展的线性单链分枝。这种分子状态在噬菌体和细菌中相当罕见。在腺病毒复制中,这样的中间体却构成了大量的胞内 DNA(图 21-47)。

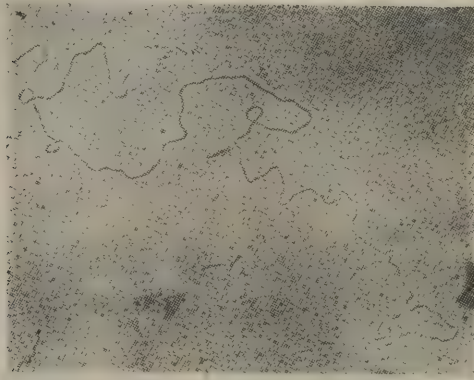
图 21-48 是目前普遍接受的腺病毒 DNA 复制模型。有两种合成途径,在主要途径中,合成和置换完成后(经过 I 型中间体,如图),产生一个双链子分子,同时释放一个母体单链。后者作为模板,合成一互补链。于是,生成两个半保留的复制分子。在次要途径中,经过互补合成,从释放链的 3' 端开始,生成 II 型中间物,它在电镜下观察为一双链片段连接着一条长的单链。这一中间体继续生长,成为完全的双链分子。这个是次要途径,因为起始作用是从分子的两端偶然地引发的。这样的双起始分子,当两个相对运动的复制叉相遇时,母链不能再靠碱基对而维系在一起,而是相互分开,形成 II 型分子。它的复制通过简单延伸而完成,这一事件称为复制叉湮灭(fork annihilation)。不久还将讨论图 21-48 中括号内的形式。

腺病毒 DNA 链的延长机理不同于我们至今所考虑过的其他双链噬菌体和病毒体系的机理。腺病毒 DNA 的 l 链和 r 链分开来复制,各自沿着适合于 DNA 聚合酶活力的方向进行复制。这样,正如这个模型所设想的那样,并不进行不连续复制,即不需要前体片段。事实上,有相当的实验支持这种看法,即两条链都进行连续复制。

出于两个原因,人们对这种复制的起始给予非常的重视:(1) 在末端的起始带来了与大肠杆菌噬菌体 T7 所面临过的同类的问题(见第十五章);(2) 没有使用 RNA 引



(a)



(b)

图 21-47 DNA 复制的电镜照片。(a) 类型 I 分子。在平行的双螺旋中其一条单链分枝正在被取代。(b) 类型 II 分子。完全被取代下来的单链又正在转变成双链 DNA；箭头指出生长中的 3' 末端。

物。而且,如模型所示有两类起始过程: 发生在双链 DNA 分子末端(产生类型 I 分子)和发生在单链 3' 端(产生类型 II 分子)。也有人认为各种不同的起始方式可能更多的是表面现象,而不是实际事实。由于末端反向重复,所以它的单链可以形成末端双链区,如图 21-48 括号中所示,这些区域称为“扳手”。这个形式有一个末端双链区,与在线性双链腺病毒 DNA 中所发现的一样。因此起始一个“扳手”也与起始病毒的 DNA 的复制相同。目前,尽管“扳手”分子可以在实验室中经连续变性和复性而得到,但在被感染的细胞中尚未分离到,因此还不知道在生物体内这种分子是否真正存在。

引发是通过衍生出一种 55kd 端基蛋白质完成的,这个蛋白质与非生长的病毒 DNA 相连接。55kd 的端基蛋白质是由一个编码出分子量为 80 000 的 80kd 蛋白的病毒基因合成出来的。生长链的 5' 端共价地连接在 80kd 蛋白质上,在图 21-49 表示的机理中,这个 80kd 的蛋白质用于起始。原初起始事件中, dCTP 的  $\alpha$ -磷酸基(连到脱氧核糖上的磷酸基)与 80kd 蛋白质分子中的丝氨酸的  $\beta$ -OH 基因之间形成酯键。酯键的形成伴随有 80kd 蛋白质和 DNA 亲代链末端的缩合作用。这种缩合作用可能因 80kd 蛋白质结合到模板链中的特殊碱基序列而受到促进,也可能通过 80kd 蛋白质结合到处于另一条模板链上的 55kd 蛋白质而受到促进。看来,后一种可能性是正确的。因为从



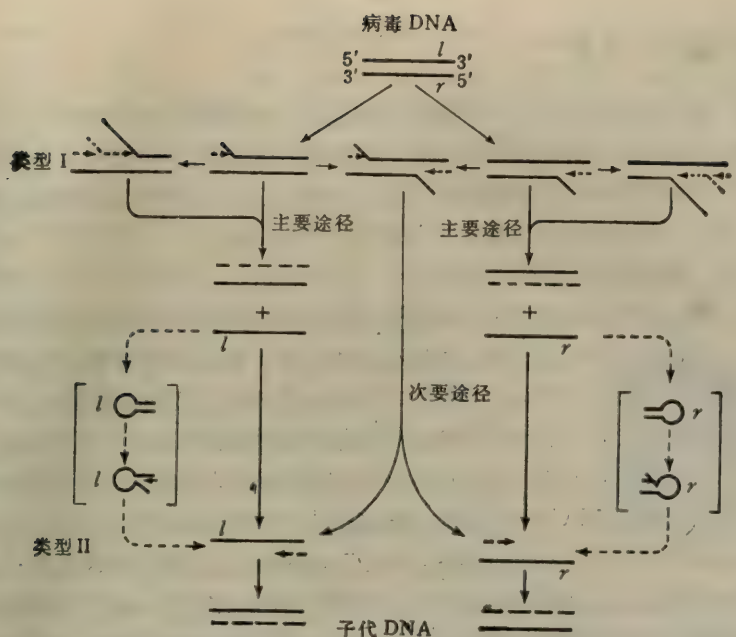


图 21-48 腺病毒 DNA 的复制图。粗黑色箭头表示主要的途径；细黑色箭头表示次要的途径；虚线箭头表示假设的反应。

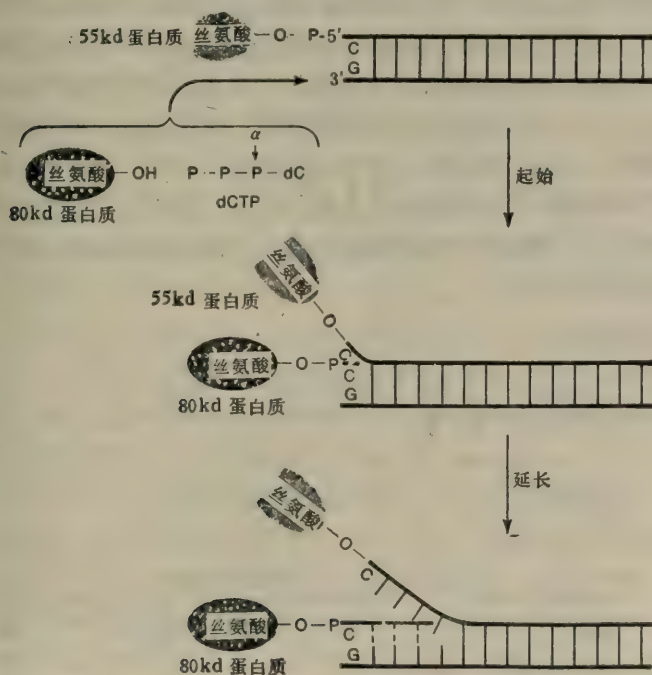


图 21-49 在腺病毒 DNA 复制的起始和延长中终端蛋白质的功能。在早期识别阶段可能通过连着 dCTP 的 80kd 蛋白质把它结合到 55kd 蛋白质上。

体外实验看,不带有 55kd 蛋白质的 DNA 作为模板对于复制的起始来说太弱。子链的生长从 dCTP 引物的 3'-OH 开始。在晚些时候,一个病毒蛋白质切割 80kd 蛋白质,使其成为 55kd 蛋白质。

## 腺病毒 DNA 的转录

腺病毒 DNA 的转录方式非常复杂,不仅有早、晚类型的 mRNA,而且与迄今已讨论过的小病毒不同,它至少有五个早期转录单位,每一个单位都从一个专一的启动子起始。这五个单位不是同时被活化,全部起始长达 3 小时,而且有些单位受它们的基因产物的负调节作用。最后,每一个早期 mRNA 在翻译之前都经过一个拼接过程(事实上, mRNA 的拼接最早是在腺病毒的研究中发现的)。晚期 mRNA 的形成方式更为复杂。关于这种复杂性的一个例子是晚期 mRNA 的合成需要有病毒 DNA 复制作用的开始。与噬菌体 T4 的情况不同,它只需要复制的起始,而不是复制本身,因为如果复制在开始后不久就被抑制,仍能继续合成晚期 mRNA。实际上,晚期 mRNA 能连续不断地起始和形成,但在 DNA 复制开始之前,这种转录作用便会在起始密码子之后约 5 000 个碱基的地方终止。晚期 mRNA 还受到广泛的加工,形成许多不同的 mRNA 分子。这个非常复杂的加工作用是用来说明从一段 DNA 中得到最大信息的一个绝妙的例子。在这一段 DNA 中,有一个晚期启动子和一个终止信号,然而,却在 mRNA 上面有四个不同的、可以切割并且可加上多聚 A 尾的位点。五个片段的每一个都有 AUG 密码子,故可以被翻译。然而,每个片段都包含着几个毗邻的内含子和几对切点,于是可以进行剪接。图 21-50 给出一个例子,说明不同的拼接可以产生不同 mRNA 分子。通过这个多重拼接方式的作用,在一个启动子控制下可以合成出至少 14 种晚期蛋白质。注意,每一次晚期转录起始后,只生成一个 mRNA 分子。若不经加工,只有从最靠近 5' 端 AUG 密码子的序列才能翻译出蛋白质。由于其他的每一种蛋白质的合成需要有一次或多次剪接过程,所以这些蛋白质的合成速率或是其中每一种晚期蛋白质的最终浓度,都是由每个剪接事件的过程以及添加多聚 A 尾的概率所决定的。因为每一事件的概率不尽相同,所以得到不同浓度的蛋白质。这种翻译作用的调控类型在真核体系中是常见的,但不存在于原核体系中。在原核中,多顺反子 mRNA 中不同顺反子的翻译概率取决于核糖体结合位点的强度和起始密码子到后面终止密码子的远近。

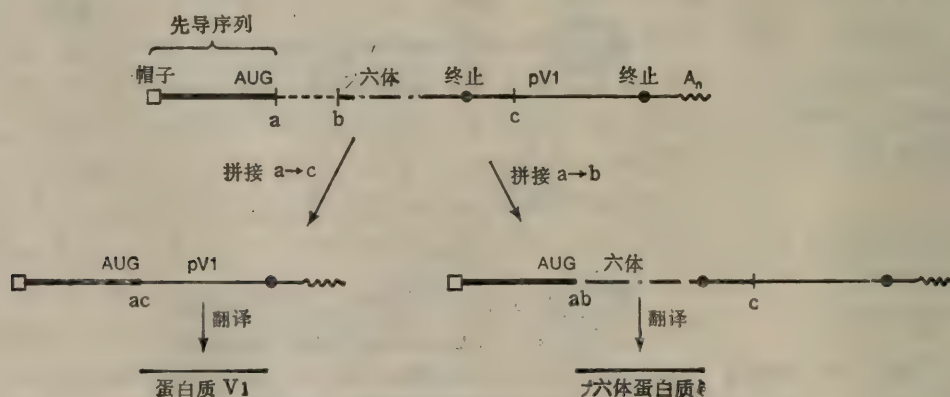


图 21-50 腺病毒晚期 mRNA 形成过程中使用的一种类型的剪接。

## 疱疹病毒

疱疹病毒是一种很大的有被膜的病毒,含有 100 多个基因,它们在两个方面有别于其



牠的 DNA 病毒: (1) 子代片段在病毒颗粒中以某种方式重新装配成病毒颗粒; (2) 它们具有三类 mRNA, 称为  $\alpha$ 、 $\beta$  和  $\gamma$ , 每一类合成的时间独立于 DNA 的复制过程。初步数据表明,  $\alpha$ -mRNA 分子所翻译出的基因产物激发  $\beta$ -mRNA 的转录,  $\beta$ -mRNA 的基因产物又启动  $\gamma$ -mRNA 的转录。正如所有的病毒中的情况, 转录后的 mRNA 前体还要再经过广泛的拼接, 才能形成有功能的 mRNA 分子。

## 牛痘病毒

痘病毒(牛痘病毒是它的一个例子)是已知的最大病毒, 有些含有 200 多个基因(图 21-51)。正像非常大的噬菌体一样, 痘病毒合成出大量用于 DNA 合成和蛋白质合成的酶和因子; 它们可以通过关闭宿主中的大分子合成而从中获利。事实上, 正是这一关闭最终导致宿主细胞的破裂和子代病毒粒子的释放。引起人们特殊兴趣的是牛痘病毒 DNA 复制的机理。牛痘病毒 DNA 分子是一个连续环状单链, 几乎全部碱基都配对, 从而形成一个双链的闭合发夹(见图 21-51)。换一种方式讲, 可以认为形成一个末端封闭的双链线性分子。引发在复制起始点发生, 然后双螺旋打开, 使得先导链前进。另外又合成出前体片段来作为滞后链的成分。当分子几乎全部被复制出来时, 起始点处产生一个单链缺刻。使得亲代链与子代链的杂合体分子上产生一个双链的缺刻。在拷贝最后一个碱基对时, 在正对着起始点的位置上造成另一个缺刻, 由此产生两个完整的双链分子。然而, 这些分子是如何成为双链闭合发夹形式的, 目前还不清楚。

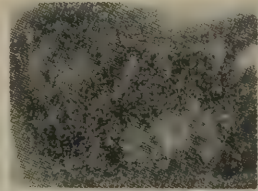


图 21-51 牛痘病毒的电镜照片。

## 关联病毒

在许多 DNA 和 RNA 病毒实验中观察到, 从被感染细胞培养体系中得到的子代病毒, 样品中含有不同大小的两种颗粒。用速度沉降 (velocity sedimentation) 或平衡 (equilibrium) 法都可以将它们分开。人们发现两类中较小的那种是非感染性的, 称为缺陷型颗粒。也许有人奇怪, 既然它们是没有感染性的, 那么它们是如何混入到新鲜繁殖的病毒样品中的呢? 事实上, 这些缺陷型病毒只有在被大颗粒病毒侵染过的细胞再进行感染之后, 才能够繁殖。这种缺陷型病毒称为关联病毒 (associated virus)。在这种病毒中, 研究最为深入的有含 RNA 的劳氏肉瘤关联病毒 (Rous-associated virus, RAV) 和含 DNA 的腺关联病毒 (adenoassociated virus, AAV)。关联病毒有两类: 其中一些是大颗粒病毒的缺失突变型; 而其他一些则是完全不同的病毒。在以上两种情况中, 它们的缺陷都被有侵染性的病毒颗粒的病毒基因所补足。这一现象已经偶尔地在细菌的噬菌体体系中见到。比如, 大肠杆菌的噬菌体 P4, 它缺少许多必需基因。但是如果先用噬菌体 P2 去感染宿主, 那么 P4 就能很好地生长在也受 P2 感染的宿主细胞中。因此, P4 就相当于病毒中的关联病毒, 它需要通常被称为助手 (helper) 的 P2 的协助, 才能完成其生活周期。

## 感 染 性 DNA

在一定条件下,噬菌体 DNA 能感染有些细菌,从而引动一个感染周期。这个过程往往要求在加入感染性 DNA 之前,先用溶菌酶处理细菌,使之成为原生质球,或者将感染性 DNA 加入到悬浮在  $\text{CaCl}_2$  溶液中的细胞外(见第二十章)。

许多 DNA 病毒的纯净核酸也具有感染性,并且细胞通常不需要经过任何预处理或处于任何特殊的溶液中就能吸收病毒 DNA。负链病毒 RNA 本身是非感染性的,因为它不能生成复制酶或逆转录酶。这些必要的酶是宿主中所没有的,通常应由病毒粒子带入宿主细胞。正链病毒 RNA 则常常是有感染性的,植物病毒 RNA 同样也是有感染性的。

纯化的病毒核酸引动感染周期的能力在实验病毒学中十分重要。由于自然界中绝大多数动物病毒频频地进行着遗传重组,所以许多年来很难选育出像噬菌体研究中的那样一些相当有价值的多重性突变体 (multiple mutant)。然而,运用重组 DNA 技术,用直接的方式就可以获得 DNA 病毒的多重性突变体。首先,两个突变分子经同一种限制性内切酶作用,接着纯化所得到的片段,然后再将适当的片段进行混合,这样就可以形成一个含有两种突变的分子。最后再用纯化了了的这种 DNA 分子去感染细胞,终于产生出具有双突变的病毒。

## 干 扰 素

干扰素是一种糖蛋白,所有的脊椎动物细胞在对 RNA 病毒的感染作出反应时合成并分泌干扰素。分泌出的干扰素可以使未受侵染的细胞在感染中暂时性地活下来。那些暂时不受病毒感染之害的细胞处于抗病毒状态。虽然只有 RNA 病毒才诱导产生干扰素,但是对于受病毒感染的细胞来说,干扰素对于所有的动物病毒(包括 DNA 病毒及 RNA 病毒两类)造成的感染都有抑制作用。

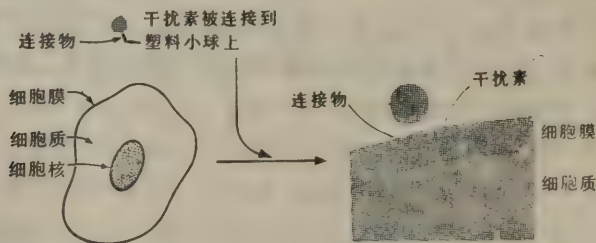


图 21-52 干扰素在细胞膜中作用的一个试验。干扰素分子通过一个短的( $10-20 \text{ \AA}$ ) 疏水连接分子(称为连接物)连到一个在显微镜下可以见到的塑料小球上。这个连接物比细胞膜的厚度还小一些。当实验者把这个单位加到细胞中后,细胞就达到抗病毒的状态,虽然塑料小球不能进入细胞膜。

很小浓度的干扰素( $10^{-11} \text{ mol/L}$  或  $10^{-9} \text{ g/cm}^3$ ) 就能产生抗病毒状态。干扰素即使与一个不能穿过细胞膜的人工合成的塑料小球共价相连,也仍具有活性,这意味着干扰素作用于细胞表面。其他实验表明干扰素与细胞外膜紧密结合,说明这个膜是干扰素的作用



用位点。

一般认为干扰素是多数脊椎动物抵御 RNA 病毒的第一道防线, 因为被 RNA 病毒感染的细胞产生干扰素所需时间比被感染动物的免疫系统产生抗体所需要的时间短得多。这个概念是由下面的实验结果得来的, 该实验已用很多种不同的形式进行过。用一种可在细胞培养的小鼠细胞中生长的病毒去感染小鼠, 一般不会引起明显的症状。假设受感染的小鼠在被感染期间产生干扰素, 如果在感染的同时注射抗小鼠干扰素的抗体(单独地从兔或山羊中制备), 小鼠得不到防护便会得重病甚至死亡, 因为病毒大量生长而造成细胞损伤。因此认为只有依靠小鼠中产生的干扰素(是对 RNA 病毒感染作出反应而产生的)发挥其抵御活性, 小鼠才能活下去; 不然, 病毒侵入机体后就会造成感染, 甚至死亡。因为对于小鼠来说, 等到病毒感染几天后它才产生出抗体则是太晚, 实际上抗体并不能抑制住病毒的感染。

干扰素的研究存在两个问题: 一是鉴定出在病毒生活周期中能使被感染的细胞合成出干扰素的那种物质的组分; 二是干扰素对后续的病毒合成进行抑制的机理。鉴定上述诱导物质的线索是干扰素几乎只由 RNA 病毒诱导。更进一步看到, 除非是双链 RNA 病毒感染, 一般地说在复制前从来不会发生干扰素的合成。这两个发现告诉我们, 双链 RNA 可能是合成干扰素的诱导者。这个概念已经通过以下方法测验过: 即向未被感染的细胞加各种人工合成多聚核苷酸, 结果发现单链多聚核苷酸如 poly C 和 poly U 是很弱的诱导者, 而双链多聚核苷酸确能有效地诱导干扰素的合成。目前知道的最有效的诱导剂是 poly I · poly C, 其中 I 是次黄嘌呤。双链 RNA 是唯一的诱导者的概念不是很严格的, 因为单链 RNA 也能诱导干扰素的合成, 虽然效率很低。现在, 人们一般都认为 RNA 病毒感染的细胞中存在的双链 RNA 很可能是诱导者(应该记得, 双链 RNA 并不是复制形式, 而可能只是 RNA 病毒生命周期中一个无用的副产品)。

干扰素的抗病毒效应在于它抑制了翻译过程。下面将对三种独立的抑制途径(见图 21-53)进行讨论。

1. 第一条途径包含一种弱化抑制作用。这是一种待命状态, 它通过一种完全不清楚的方式造成 tRNA 的不足, 可能使细胞变得衰弱, 当然也就减小了病毒感染的效率。但是, 如果所有暴露于病毒的细胞都被如此地弱化, 至少在一段时间内也将对动物不利。然而由于暴露的细胞又是病毒感染的直接环境, 所以弱化这些细胞对动物来说却又有救活其整个机体的价值。这一条途径很可能也能用来说明 DNA 病毒的感染受到弱化的抑制作用。

第二和第三条途径直接抑制病毒感染。在没有被感染的、已用干扰素处理的细胞中, 这两条途径处于无活性状态, 当加入双链 RNA 时, 它们就被激活了。

2. 在第二条途径中, 一种能分解病毒 mRNA 的 RNA 酶被激活, 这个激活作用可通过一个体外实验来说明: 向已用干扰素处理过的细胞提取物中加入高浓度的双链 RNA (双链 DNA 和 DNA-RNA 杂合子都不行)。于是一种合成酶(目前还未被提纯)就将三个分子的 ATP 连接成寡聚核苷酸 pppA2' p5' A2' p5' A (简写作 2,5A, 见图 21-53)。新生成的这个寡聚核苷酸再去激活早就存在的、没有活性的宿主 RNA 核酸水解酶, 然后这种酶又将 mRNA 降解成核苷酸, 这样就很快地终止了感染周期。

3. 当双链 RNA 浓度很低的时候则发生第三条途径。这时, 宿主中早已存在的一种

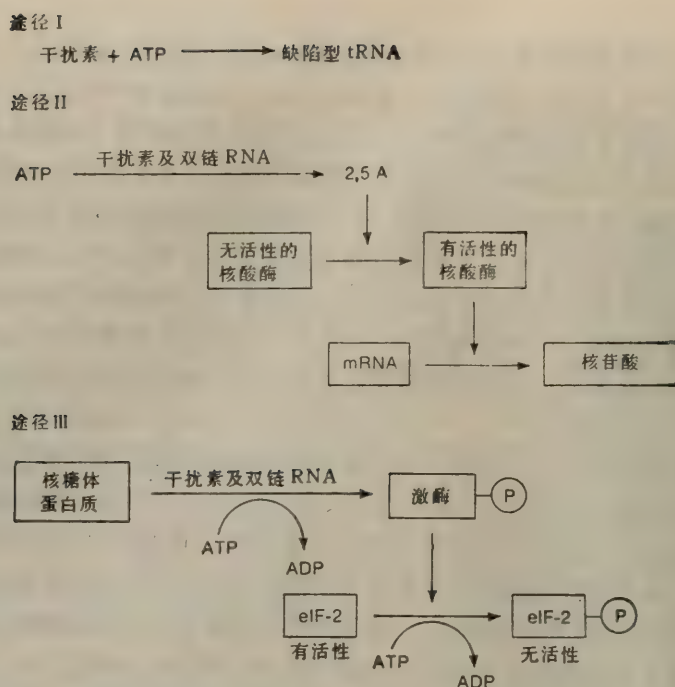


图 21-53 干扰素抑制翻译过程的二种途径。途径 II 及途径 III 是主要的途径。

核糖体蛋白质经磷酸化作用转变成蛋白激酶。这种蛋白激酶能使有活性的延长因子 eIF-2 受到磷酸化作用而失去活性，从而抑制所有的蛋白质合成。

被感染细胞产生的干扰素是很少的，以致研究工作者无法获得足够的纯的干扰素，使干扰素的研究长期受到阻碍。当人们把几种不同的干扰素基因通过 DNA 重组技术克隆到大肠杆菌中后，才克服了这个困难，不久，就有大量的纯干扰素供应。大量地供应纯干扰素使人们把它作为临床上的抗病毒药剂的希望变成可能。人们也认为，当肿瘤是由于病毒引起的时候，干扰素可能有抗肿瘤活性。

## 植 物 病 毒

植物病毒都是 RNA 病毒。对它们的大量研究落后于对动物病毒的研究，首先是因为在植物中为了起始病毒的发育需投入的感染量必须极大；而且，还不得不用生长很缓慢的植物作为宿主。在最近几年中，已有几种方法使病毒能在植物细胞的原生质体（除去了细胞壁的细胞）中增殖，这样逐渐地收集到了一些关于植物病毒生活周期的详细材料。大多数植物病毒的基因产物的分子很大，其分子量超过了病毒 RNA（如果不重叠）应编码的信息容量。不过这个问题可以通过事实上存在着重叠基因来解释清楚。目前对烟草花叶病毒 (TMV) 研究很透彻，在图 21-54 中给出了 TMV 的基因在其 RNA 上的排列。据认为病毒蛋白 VP1 和 VP2 是由同一个 mRNA 分子翻译并且不再经切割而产生的。但是，在体外从来没有实验证明过不存在多聚蛋白。在体外，从 RNA 病毒粒子不能直接地翻译出外壳蛋白 (coat protein)，外壳蛋白只在感染后由一个 mRNA 翻译出来，这个 mRNA 与



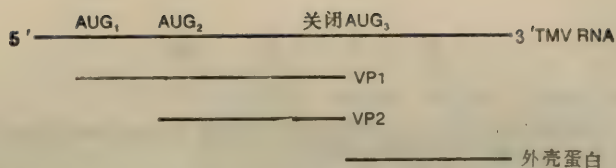


图 21-54 从烟草花叶病毒 (TMV) 翻译的三种蛋白质。

病毒颗粒的 RNA 有同样的长度,但形状不同。据认为,新合成但还未包装的 TMV RNA 有一个开放的核糖体结合位点,但在 RNA 病毒颗粒中,这个位点被关闭。如果是这样,外壳蛋白的合成受到的调节就像大肠杆菌噬菌体 M13 的外壳蛋白的合成调节一样。

植物病毒不诱导细胞裂解,结果在植物细胞中能积累巨大数量的颗粒多至  $10^7$  个(见图 21-55)。敏感的植物细胞继续受感染是病毒颗粒通过细胞膜上的孔洞和连接管在细胞间进行直接运输的结果。有时颗粒浓度甚高,高达可在细胞中形成结晶,使得 X 射线衍射家能对病毒结构做详细的研究。由于 TMV 容易纯化,在其结构研究中已经做了大量工作\*。在病毒结构研究中给人印象最深刻的是 1955 年 Heinz Fraenkel-Conrat 和 Robleys Williams 的工作,他们揭示出纯化的 RNA 和外壳蛋白分子可以重组成有感染能力的 TMV 病毒颗粒。通过 X 射线衍射研究了各种中间形式,依次的过程已在第六章说明。

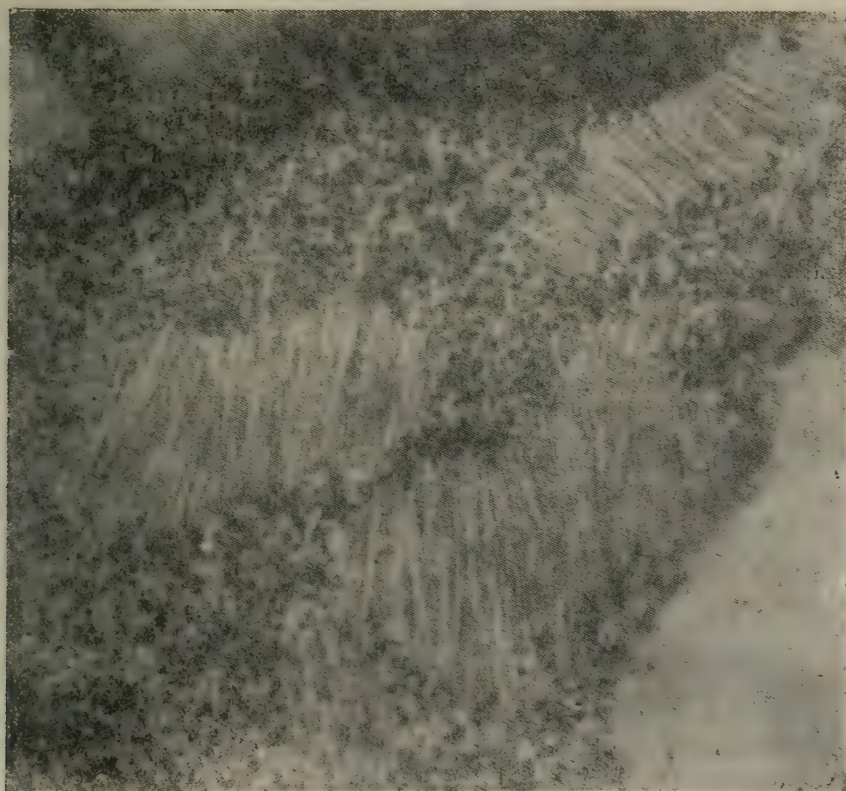


图 21-55 烟草叶子的横截面,示出了 TMV 颗粒的排列。

\* 读者也许还记得“一个碱基的改变足以引起一个氨基酸发生变化”的第一个工作就是在以 TMV 的外壳蛋白为材料的研究中做出来的。

## 参 考 文 献

- Bishop, J. M. 1978. "Retroviruses." *Ann. Rev. Biochem.*, 47, 35—88.
- Bishop, J. M. 1980. "The molecular biology of RNA tumor viruses: a physician's guide." *New England J. of Med.*, 303, 675—682.
- Botchan, M., J. Stringer, T. Michison, and J. Sambrook, 1980. "Integration and excision of SV40 DNA from the chromosome of a transformed cell." *Cell*, 20, 143—152.
- Butler, P. J. T. Finch, and D. Zimmer. 1977. "Configuration of tobacco mosaic virus during virus assembly." *Nature*, 265, 217—219.
- Butler, P. J. G., and A. Klug. 1978. "The assembly of a virus." *Scient. Amer.*, November, pp. 62—69.
- Caspar, D. L. D., and A. Klug. 1962. "Physical principles in the construction of regular viruses." *Cold Spring Harb. Symp. Quant. Biol.*, 27, 1—24.
- Cold Spring Harbor Laboratory. 1979. *Viral Oncogenesis. Cold Spring Harb. Symp. Quant. Biol.*, Vol. 44.
- Crawford, L. V. 1980. "The T antigens of simian virus 40 and polyoma viruses." *Trends Biochem. Sci.*, 5, 39—42.
- Diener, T. O. 1979. *Viroids and Viroid Diseases*. Wiley.
- Diener, T. O. 1981. "Viroids." *Scient. Amer.*, January, pp. 66—73.
- Eckhardt, W. 1977. "Genetics of polyoma virus and simian virus 40." *Comprehensive Virol.*, 9, 1—26.
- Eckhardt, W. 1981. "Polyoma T antigens." *Adv. Cancer Res.*, 35, 1—26.
- Fenner, F., B. R. MacAulian, C. A. Mims, J. Sambrook, and D. O. White. 1974. *The Biology of Animal Viruses*. Academic Press.
- Friedmann, R. M. 1977. "Antiviral activity of interferon." *Bacter. Rev.*, 41, 543—567.
- Graessman, A., M. Graessman, and C. Mueller, 1981. "Regulation of SV40 gene expression." *Adv. Cancer Res.*, 35, 111—151.
- Henle, W., G. Henle, and E. T. Lennette. 1979. "The Epstein-Barr virus." *Scient. Amer.*, July, pp. 48—59.
- Holmes, K. C. 1980. "Protein-RNA interactions during the assembly of tobacco mosaic virus." *Trends Biochem. Sci.*, 5, 4—7.
- Hynes, R. 1980. "Cellular location of viral transformation proteins." *Cell*, 21, 601—602.
- Joklik, W. K. (ed.). *Principles of Animal Virology*. Appleton-Century-Crofts.
- Kitamura, N., B. L. Semler, R. G. Rothberg, G. R. Larsen, C. J. Adler, A. J. Dorner, E. A. Emini, R. Hanecak, J. J. Lee, S. van der Werf, C. W. Anderson, and E. Wimmer, 1981. "Primary structure, gene organization, and polypeptide expression of poliovirus RNA." *Nature*, 291, 547—553.
- Knight, C. A. 1975. *Chemistry of Viruses*. Springer-Verlag.
- Lebeurier, G., A. Nicolaieff, and K. E. Rischards. 1977. "Inside-out model for self-assembly of tobacco mosaic virus." *Proc. Nat. Acad. Sci.*, 74, 149—153.
- Lebowitz, P., and S. Weissman, 1979. "Organization and transcription of the simian virus 40 genome." *Current Topics in Microbiol. and Immunol.*, 87, 43—172.
- Luria, S. E., J. E. Darnell, D. Baltimore, and A. Campbell. 1978. *General Virology*. Wiley.
- Mitra, S. 1980. "DNA replication in viruses." *Ann. Rev. Genes.*, 14, 347—398.
- Nathans, D. 1979. "Restriction endonucleases, simian virus 40, and the new genetics." *Science*, 903—909.
- Shimotono, K., S. Mizutani, and H. M. Temin. 1980. "Sequence of retro provirus resembles that of bacterial transposable elements." *Nature*, 285, 550—554.
- Silverstein, S. G., J. K. Christman, and G. Acs. 1976. "The reovirus replicative cycle." *Ann. Rev. Biochem.*, 45, 375—408.
- Temin, H. M. 1972. "RNA-directed DNA synthesis." *Scient. Amer.*, January, pp. 24—33.
- Temin, H. M. 1980. "Origin of retroviruses from cellular movable genetic elements." *Cell*, 21, 599—600.
- Tooez, J. 1981. *The Molecular Biology of Tumor Viruses*. Cold Spring Harbor Lab.
- Weinberg, R. A. 1980. "Integrated genome of avian viruses." *Ann. Rev. Biochem.*, 49, 197—226.
- Williams, R. C. and H. W. Fisher, 1974. *An Electron Micrographic Atlas of Viruses*. Thomas.
- Ziff, E. B. 1980. "Transcription and RNA processing by the DNA tumor viruses." *Nature*, 287, 491—499.



## 第二十二章 肿瘤病毒和癌基因

1911年 Peyton Rous 证明,在小鸡中有一种病毒可以诱发产生肉瘤(一种肿瘤),这种病毒现在称为劳氏肉瘤病毒(Rous sarcoma virus)。从那以后,人们发现许多病毒都能使动物致癌,但在人类中,致癌病毒(oncogenic viruses)并不多见。几乎所有的DNA病毒系都有致癌病毒。可是在RNA病毒中,只有某些逆转录病毒能致癌。致癌的DNA病毒和RNA病毒在以下几方面有着明显的不同:

1. 对两种细胞的影响不同:某特定的病毒(指DNA病毒)可以对某种细胞作用,即它可对受纳细胞(permissive cell)作用,在细胞中大量地生长繁殖,产生大量的子代病毒;而对另一种细胞,即非受纳细胞(non-permissive cells),则无法引起产生大量子代病毒,而只能引起肿瘤。然而,逆病毒则可以引起受纳细胞产生肿瘤。

2. DNA 致癌病毒不能使大部分受纳细胞受到感染,只有小部分细胞能够受病毒感染并转变为肿瘤细胞。而RNA 致癌病毒(逆病毒)则可使允许它生长的大部分细胞(即受纳细胞)都发生癌变。

3. DNA 病毒和逆病毒都可将其DNA整合到宿主的染色体中,但是形成前病毒(provirus)的相对或然率(relative probability)即可能性很不相同。在逆转录病毒的生活周期中,整合是必不可少的(见第二十一章真核病毒)。而DNA病毒则很少进行整合,因为其繁殖并不依赖于整合。在下面我们还要讲到整合。

病毒致癌研究中的第一个重大突破是发现了转化细胞(transforming cell)。转化细胞与正常细胞在各种性质上都不相同,转化细胞在被注射到适当的动物中后常常可以致癌。因此在这一章我们首先讲讲转化细胞。

### 转 化 细 胞

除了一些特殊的血细胞和免疫系统的细胞以外,动物细胞在培养基中只生长在培养瓶的表面上。如果把一片典型的组织放在玻璃的或经过特别处理的塑料的表面上,这片组织会长出各种类型的成纤维细胞。另外,用胰蛋白酶可将该组织分散为单个的细胞,在适当的培养基中,悬浮的成纤维细胞会吸附到容器的表面上,并开始生长(大多数其他的细胞基本上不生长)。这种组织培养物叫做原初培养物(primary culture)。其特点为细胞只分裂几代以后大多数就死去,这叫做危机(crisis)。但是,如果把剩下的几个细胞在培养基中再放置几个月(在此期间,细胞几乎不或完全不分裂),某些细胞又开始生长,最终形成一个具有生长能力的细胞培养物。这种细胞培养物叫做确立细胞株(established cell line),它与组织培养中的原初细胞大不相同,最显著的是原初细胞在几代之后便死亡,而这些细胞能无限地生长。从这种意义上来说,他们是“不朽”的,很显然,这种细胞是不正常的;然而,在大多数的细胞生长研究中,人们用的都是这种细胞。

如果把典型的确立细胞株放在培养瓶中,细胞会不断地生长和分裂,直到形成连续的

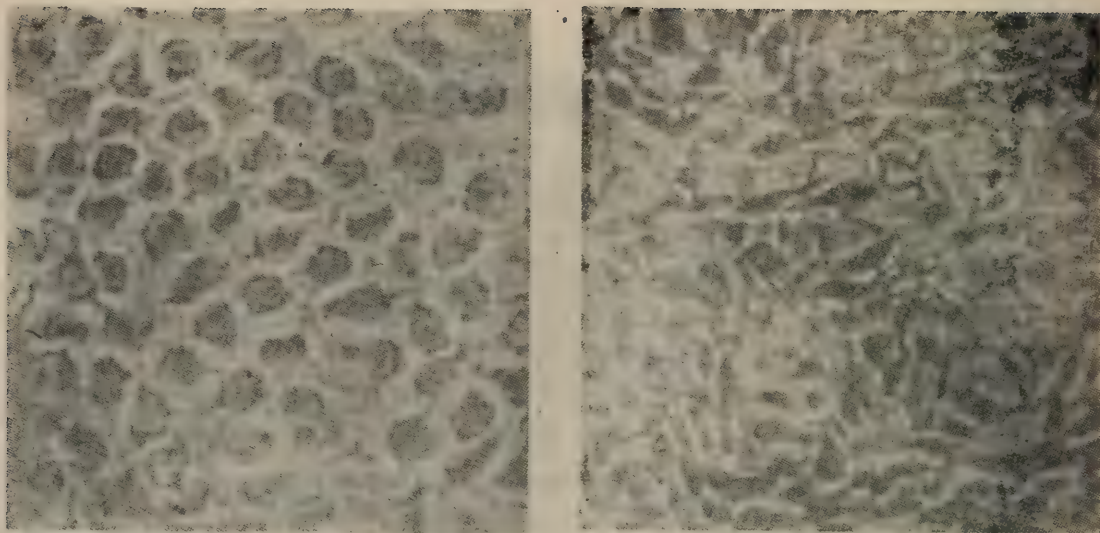


图 22-1 (a) 正常的小白鼠 3T3 细胞的汇合层。(b) 被多瘤病毒转化的 3T3 细胞,不仅形状不同,而且排列比正常细胞混乱。

单细胞层,生长才停止,这种现象称为密度依赖性生长 (density-dependent growth), 以前称为接触抑制 (contact inhibition)。生长停止时,大多数细胞是有序排列的 (图 22-1 (a))。用胰蛋白酶可把这些连绵的、相互汇合的细胞从表面上洗下来,如果再把这些细胞放在新的表面上,并提供新的培养液,它们又能分裂、生长,再度形成连绵的单层细胞。

如果一种确立细胞株不允许某特定病毒生长(此确立细胞株就是非受纳细胞,它可以被转化),当它的原代细胞培养物被该种病毒感染后,这种细胞培养物继续培养的结果就与那些未受过感染的细胞培养物培养后的结果全然不同:当细胞生长到形成汇合的单细胞层时,有些细胞生长排列不规则了,无序生长的细胞会继续增殖,最终形成一个混乱的多层细胞团。如果用胰蛋白酶将这些细胞洗下并分散开来再继续培养,这时所形成的细胞培养物表现出无序的多层生长[图 22-1(b)]。当细胞具有这种特性时,我们就说它们是被转化了,称做转化细胞。转化也可由其他方式进行。例如,接受辐射或化学致癌物处理(后面将要谈到),或甚至强迫细胞在高浓度\*的条件下生长许多世代也能造成转化。

与正常细胞相比,转化细胞还有几个不同之处:

1. 可在悬液中生长(而不是只能在表面上生长)。
2. 它们需要生长在含有较低浓度血清的培养基中。
3. 细胞表面有所改变,细胞膜更具有流动性。
4. 染色体数总多于正常的双倍体数,可能为三倍体、四倍体,或个别的染色体具有额外的拷贝数。

但是,它们最显著的特征在于,如果将这些转化细胞注射到同一种属的动物中(如转

\* 这种差别在浓度上可能是很小的。例如,在小鼠细胞系建立的早期阶段,细胞浓度维持在  $3 \times 10^5$  个细胞/毫升,这是一个正常的细胞系;但如果使细胞浓度维持在  $1.2 \times 10^6$  个细胞/毫升的条件下(正好是 4 倍,处于波挤压的状态),就可由转化细胞组成一个确立细胞株。



化的小鼠细胞注入小鼠),常常可以致癌。转化细胞内部都有一个或更多个整合了的病毒 DNA, 这是病毒的转化细胞的一个共性,也为我们研究转化的分子基础提供了线索。

### 检测整合的病毒 DNA

在研究病毒诱变的转化系统中,人们不仅可以证明转化细胞中有整合的病毒 DNA, 而且还能测定该整合序列的数目和位置。在这一部分我们将主要介绍这些测定方法。

首先指出病毒转化细胞中含有病毒 DNA 的是杂交实验。在体外,将纯化的多瘤病毒 DNA 转录为 RNA,再用<sup>3</sup>H标记该 RNA。从多瘤病毒转化细胞中提取得到 DNA 可以与这种标记 RNA 杂交,而未被转化细胞的 DNA 则不行。病毒 DNA 可以整合到宿主 DNA 中,也可以以质粒形式存在(如第十六章所讲到的两种溶源性噬菌体)。用一个简单的实验即可区分这两种情况。从被多瘤病毒转化的细胞中提取总 DNA,将 DNA 随机地打断为各种分子量的片段,通常这些片段都大于多瘤病毒 DNA (分子量是  $4.8 \times 10^6$ )。用区带离心将这些片段分离,然后使之变性并和<sup>3</sup>H 标记的多瘤病毒 RNA 杂交,从而测定是否含有多瘤病毒 DNA。如果病毒 DNA 以质粒形式存在,则所有的多瘤病毒 DNA 将以分子量为  $4.8 \times 10^6$  的单一组分沉降,但如果该 DNA 是整合到宿主染色体中,而且细胞的总 DNA 又是被随机打断的,那么转化细胞含有的多瘤 DNA 就会是大小不一的片段。实验结果与后一种假设相符,<sup>3</sup>H 多瘤病毒 mRNA 可以和任意大小的 DNA 片段杂交。这个结果清楚地说明多瘤病毒 DNA 在宿主中是以整合状态存在的。

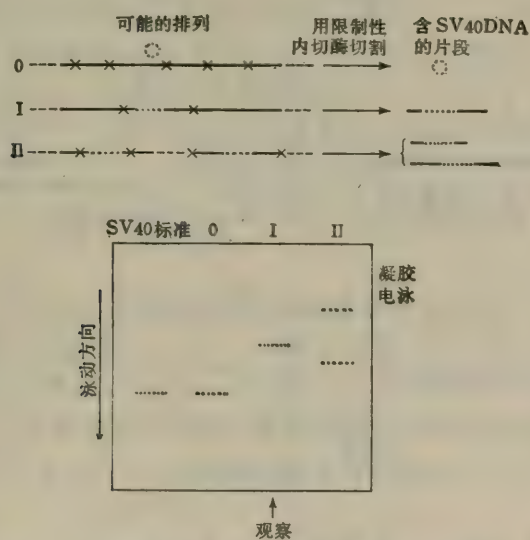


图 22-2 转化细胞中 SV40 DNA 被整合到细胞染色体 DNA 上的实验。上图黑色实线代表细胞的 DNA; 长虚线代表与此无关的大片段 DNA; SV40 DNA 用点虚线表示。“×”是限制性内切酶的作用位点。电泳图谱: 如果 SV40 DNA 以游离的质粒 (○) 形式存在, 那么将显示出与纯 SV40 DNA 有同样泳动速度的单一的条带。如果细胞 DNA 上存在着一个前病毒 (I), 那么由于这个前病毒是连到细胞 DNA 上, 所以人们可以想见电泳时条带泳动得较慢 (因为具有较高的分子量)。如果存在两个前病毒 (II), 那么电泳后会有两个条带, 因为一般限制性内切酶的作用位点分布在围绕着两个前病毒的不同地方。电泳时采用超螺旋的 SV40 DNA 作为标准, 因为它的泳动速度几乎与线性的、一个单位长度的 SV40 DNA 的相同。

后来,人们又用 SV40 进行了更为巧妙细致的研究,进一步肯定了这个结论: SV40 病毒的 SV40 DNA 是整合的。在这些实验中,人们不仅测定了前病毒的数目,还搞清了整合是否发生在特定位点上。实验方法为:用限制性内切酶将转化细胞的总 DNA 切成片段,然后进行凝胶电泳分离这些片段,并使之变性,再用  $^{32}\text{P}$  标记变性了的病毒 DNA 作为探针,进行 Southern blotting 杂交实验,鉴定该片段中是否含有病毒 DNA。

第一个实验(图 22-2)表明病毒 DNA 是整合在转化细胞染色体 DNA 的一个单一位点上的。将转化细胞的总 DNA 用限制性内切酶消化,要求这种酶在 SV40 DNA 中没有切点,而在细胞 DNA 中有许多切点。将切割后的 DNA 进行电泳。因为 SV40 既可能以质粒形式存在,也可能整合到一个或两个不同位点上,所以有两种预期的电泳结果。观察到的实际结果只与一种情况相符,只整合在一个位点上。已经从大量不同的转化细胞克隆中分离 DNA,再用这些 DNA 重复了这一实验。只在少数克隆中有两个条带,在通常情况下,电泳图谱中只有一个条带。因此可以断定,SV40 通常只

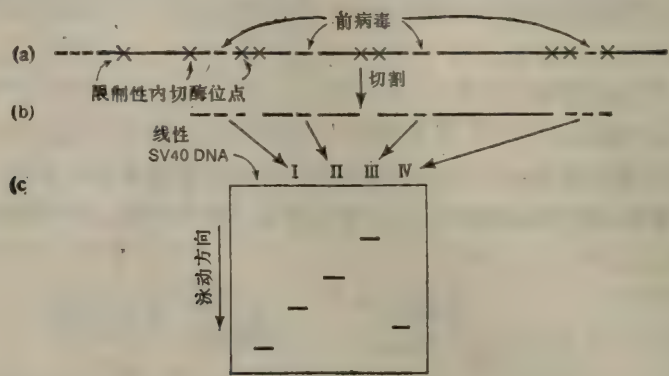


图 22-3 处于不同位置上的前病毒 DNA 电泳时在图谱上的位置。(a) 细胞 DNA 的一部分,表示在四种不同的转化细胞中前病毒整合的四种可能位置。(b) 四种不同细胞株所含前病毒 DNA 的酶解片段,由于前病毒距最近的切点位置不同,所以含前病毒的片段的大小也不同。(c) 凝胶电泳,每条带(I—IV)代表一种单独的转化细胞的克隆。

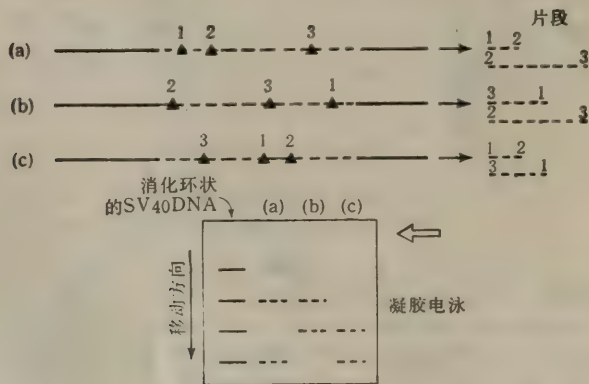


图 22-4 从三种 SV40 转化细胞株提取 DNA 后,用限制性内切酶作用片段的假设电泳图谱,假设病毒 DNA 的整合位点是变化的。如果位点是固定的,图谱将是同样的。图中没有显出含细胞 DNA 的两条带。这些分子通常比条带中 DNA 片段大得多,并且将出现在空心箭头所示的位置。



整合在一个位点上,但有些转化细胞会在不同的位点上整合两段 SV40 序列。当检验一种特殊的病毒所转化的许多个克隆时,发现各个克隆的一条电泳带或几条电泳带的位置一般都不同,在极少数情况下一个克隆的条带与其他克隆的条带相重合(如图 22-3)。这说明存在着许多潜在的整合位点,这些位点可能是随机分布的。

如果选用前病毒中有几个切点的限制性内切酶,用这种酶就可以测定整合是否发生在病毒 DNA 某一特定位点上。具体方法见图 22-4。该酶在病毒 DNA 内有切点,所以它在产生的切割片段中有一些片段小于完整的 SV40 DNA,这些片段构成一个亚组(用同一酶切割环形病毒 DNA 所得的片段则构成总的群体,亚组包括在此群体之中)。如果整合发生在一特定位点上,所有克隆都会见到同样的片段,但如果整合位点是不定的,该酶有几个切点时,电泳后就可见到( $n-1$ )个片段,表明是几个亚组。实验所观察到的结果正好与后者一致,说明 SV40 没有固定的整合位点。

很显然,SV40 DNA 可整合到宿主 DNA 的许多不同位点上。一个有趣的问题是,所有这些位点是否都在同一个染色体上呢?回答是否定的。这个答案是应用了细胞融合(cell fusion)技术得到的。在实验室中,人们常常使用极其有用的仙台病毒(sendai virus),因为它可以诱导被它感染的细胞发生融合\*。

在同种细胞或异种细胞间都可以发生融合。例如鼠和人的细胞就可以融合,所形成的融合细胞通常在最初的几次细胞分裂期间随机地丢失人的染色体,直到细胞逐步地具有完整的一套小白鼠染色体和少数几条人染色体之后形成稳定的细胞系为止。正常的鼠细胞和被 SV40 转化的人细胞可以融合,从中已分离到这样一些克隆,它们类似于鼠细胞,但又表现出转化细胞的表现型特性。因为在细胞分裂中期,鼠-人的染色体,以及人的单个的各条染色体具有不同的形态,很易分辨区分。因此,通过对于转化的鼠细胞克隆进行的检测,可以确定它含有哪一些人的染色体。经过对大量克隆进行分析,发现每个克隆中都有一条或几条不同的人染色体,这清楚地表明 SV40 DNA 可整合在许多染色体中。

## 整合的病毒 DNA 的结构

噬菌体 DNA 整合到染色体中形成前病毒,这种前病毒是噬菌体 DNA 的真实拷贝。虽然有时会有顺序改变,但是噬菌体 DNA 在整合过程中既不会有任何损失也没有任何收获。在第二十一章已经介绍,当逆转录病毒整合时,病毒碱基序列两端会失去几个碱基,而一个大片段(LTR,即长的终端重复序列)则被复制加倍。因此,逆转录 DNA 整合时,会有精巧的 DNA 重排,但不会失去编码序列。这对于逆转录病毒来说是至关重要的,因为只有整合的 DNA 才能产生子代病毒。

致癌的 DNA 病毒与此不同,整合在其生活周期中并不重要。实际上,只有在感染失败时才会发生整合。所以,当整合发生时,并不存在为保持全部编码序列的选择性压力。因为整合能力对于病毒来说并不具有使之存活的价值。实际上,人们发现病毒常常通过

\* 仙台病毒是有被膜的 DNA 病毒,它不杀死宿主,但促进宿主细胞膜的弱化过程,使两个被感染细胞能高频率地合并,从而形成一个融合细胞,或形成异核体(heterokaryon)。实验中所用的仙台病毒都是在人体中无法存活或失活的,不过这种病毒也仍能促进融合。

失去一些遗传信息而完成整合过程：DNA 片段会被删除、倒转或搅乱，这些变化的意义还不清楚，显然这些病毒的整合过程是极其复杂的。在这方面，腺病毒是最突出的一个例子。被腺病毒转化的细胞克隆总包含着整合的病毒 DNA，但是该 DNA 从来不是完整的腺病毒 DNA 分子。许多分析表明，不同的克隆具有不同的腺病毒序列。但是，有一个片段出现在所有的转化细胞中，这一发现很重要，因为它说明转化细胞的种种特性可能是由某一个或某一套特定基因造成的。

## 整合的可能后果

病毒整合可以以几种不同方式导致转化：

1. 整合可以发生在宿主的某一基因中，从而使该基因失活。如果该基因是起调节作用的，就会导致某些细胞过程的失调。
2. 整合可使某一基因与其启动子分离，从而使该基因无法表达。
3. 整合可使某一宿主基因受到病毒启动子的调控，这会使该宿主基因的产物大大过量。
4. 在前病毒状态下所表达的某病毒基因产物可以直接地或间接地引起转化。
5. 整合同时引起染色体断裂和重排，这些可能使基因表达发生改变。

以上各项中 3—5 项很重要。

肿瘤病毒与溶源性噬菌体不同，它们没有阻遏物。实际上，它们整合的 DNA 可持续地转录。而溶源性噬菌体则需要有阻遏物，以防止杀死宿主细菌的功能得到活化。阻遏物对于逆转录病毒来说也是不必要的，因为逆病毒并不杀死宿主细胞。另外，所有致癌的 DNA 病毒可以转化，也不需要阻遏物，但其原因不同。所有的致癌 DNA 病毒都有早期转录和晚期转录。这些病毒具有杀死宿主细胞的功能，但这些功能被限制在晚期才表达。关于这些我们曾介绍过。回忆一下，DNA 病毒只转化非受纳细胞，而在这些细胞中，无法进行晚期转录或晚期 RNA 的加工。

病毒不能杀死宿主细胞的事实还不足以解释伴随着转化的复杂的细胞学改变，因此，多年来人们对诱导肿瘤的病毒基因进行着很活跃的研究。下面将介绍这些基因。

## 病毒的致癌基因

在上一部分，我们已指出病毒 DNA 对宿主细胞基因表达的影响，以及致癌病毒基因的产物可以引起病毒的转化，也知道两种影响的一些实例。在这一部分，我们将研究逆转录病毒中致癌的病毒基因，它们被称为致癌基因 (oncogenes)，是引起肿瘤的一类病毒基因中的一部分。

### 蛋白质激酶和逆转录病毒引起的转化

在1970年，人们分离出了劳氏肉瘤病毒 (RSV) 的温度敏感型转化突变株。突变的病毒在允许温度和非允许温度下都能正常生长并产生子代；但只在允许的温度下有转化能力。而且，将在允许温度下转化的细胞移置到非允许温度下一两天之后它又可回复正常的细胞形态。但无论是非转化的、转化的，还是回复的细胞都可产生同样的病毒。这些实



验清楚地说明 RSV 的转化功能和复制功能是分离的。与此同时,人们还分离出一些 RSV 的非条件性缺失型突变株 (nonconditional deletion mutant),它们完全丧失了转化能力,又称作非转化性突变株 (nontransforming mutant),它们在结构上与野生型 RSV 非常类似。负责转化功能的基因叫做 src (sarcoma-producing),它与形成肉瘤的功能有关。通过比较野生型 RSV 的 RNA 和非转化突变株的 RNA,可以证明突变体是 src<sup>-</sup> 缺失突变型,它缺少了大约 1600 个碱基的 RNA。这一发现为我们研究 src 基因提供了有利的生化基础。一个令人激动的实验证实了转化完全是由 src 单独引起的。人们将分离出只包括 src 基因的限制性酶切片段。将上述 src 片段加到鸡组织培养细胞中,用一略加改变的 CaCl<sub>2</sub> 技术进行转化。结果,一些细胞发生转化。在这些细胞中,发现有整合到染色体中的 src 片段。在其他几种逆转录病毒中也发现了类似的转化基因。

src 基因最不寻常的性质很可能是仅仅一个简单的病毒基因编码的蛋白质就可以引起多种与转化表现型相关的效应。人们对此解释为:src 基因的产物是一个酶,它可修饰许多种蛋白质。

src 基因的产物为一种称之为 pp60-v-Src 的简单蛋白质。60 是指其分子量为 60000, v 表明该基因来自病毒 (我们将会发现这种命名很重要)。pp60-v-Src 是一种酶,它可将磷酸基从 ATP 上转移到细胞蛋白质的酪氨酸残基上,这种能够使蛋白质磷酸化的酶叫做蛋白激酶。大多数蛋白激酶作用在丝氨酸残基上,而 src 酶却作用在酪氨酸残基上。这很特别,因此,它又称为酪氨酸专一性蛋白激酶 (tyrosine specific protein kinase)。人们发现其他逆转录病毒的转化基因产物也是与此类似的酶。

蛋白激酶在动物中很常见,它们的主要功能是调控其他蛋白质的活性。对某些蛋白质来说,磷酸化使之活化;而对另一些蛋白质来说,则使之失活。一种蛋白激酶通常可使许多蛋白质磷酸化,因此,这些激酶的调控功能是可以进行远距离控制的 (far-reaching)。不难想像,像 src 蛋白激酶这样的酶的合成会引起细胞的生长特性发生戏剧性的变化,因为它们可以使许多酶活化或失活,而其中有些酶可能是参与生长调控的。人们对几种逆转录病毒的致癌基因的温度敏感突变株进行了研究,结果表明,酪氨酸激酶的活性与转化能力是一致的。在允许的温度下,突变株有明显的酪氨酸激酶的活性,同时也具有转化宿主细胞的能力,在转化细胞的蛋白质中也包含有磷酸化了的酪氨酸;而在非允许的温度下,既没有酪氨酸激酶活性,也没有转化能力及被磷酸化了的酪氨酸。

显然,要想了解该激酶的作用,必须鉴别出那些被磷酸化了的重要蛋白质。人们已分离出许多被 src 激酶磷酸化的蛋白质,但是大多数蛋白质的功能还不清楚。在后面我们还将介绍这些蛋白质的一些情况。

## 癌细胞中的致癌基因

因为认识到像 src 这样的一个单个的基因就可以引起转化作用,所以人们开始从癌症病人身上经外科手术切除下来的癌细胞中寻找致癌基因。前面我们提到过,src 基因可以以两种方式引起转化:将 RSV 病毒的全部 DNA 整合到一个染色体上;或只将含有 src 基因的 DNA 片段整合进去。后者使我们得到了启发:从人体癌细胞中分离出 DNA 片段,通过测定它们是否具有转化正常培养细胞的能力,就可发现癌细胞中的癌

基因。

人们首先使用人体的膀胱癌细胞系的细胞 DNA 作为实验材料,先把 DNA 切割为片段,再用这些片段去转染小鼠组织培养细胞 (NIH 3T3 细胞,这是一个确立细胞株。这类研究经常使用这种小鼠细胞株),结果小鼠细胞被转化,说明膀胱癌细胞中有致癌基因。在第二个实验中,人们把 DNA 片段分为许多个组分,然后分别测定每个组分的转化活性,结果发现只有一个 DNA 片段可以诱发转化,说明这个片段中含有人膀胱癌的致癌基因。在另一套实验中,人们用人的淋巴腺瘤细胞为实验材料,从中分离到了淋巴腺癌的致癌基因。

在发现人膀胱癌的致癌基因以后,有人提出这样一个问题:它是从哪里来的?人们给出了各种各样的理论,其中之一是突变理论,推测膀胱癌的致癌基因是由正常的基因经突变而形成的。事实的确如此,因为从正常膀胱细胞中可分离出一种 DNA 片段,这个片段的编码序列与膀胱癌致癌基因只差一个碱基:正常细胞 DNA 中的一个 G 在致癌基因中被 T 所取代。因此,人膀胱癌的致癌基因是正常基因的突变体。所以,对某些(但不是全部)其他的致癌基因来说,情况也是如此。

在下一部分,我们将会看到在正常细胞中也有 src 基因。

### src 在细胞中的对应体

人们用正常鸡细胞的总 DNA 和放射性标记的 src DNA 探针进行 Southern blotting 分子杂交,结果发现在正常细胞中有 RSV src 基因的对应体 (counterpart)。用这种方法,人们进一步在其他生物中寻找类似 src 的基因,发现在鸟类、鱼类、哺乳类和人的 DNA 中也都有 src 基因的对应体。为了区分这两种序列,人们把病毒中含 src 基因的 DNA 序列称为 v-src,而细胞中的对应序列称为 c-src。c-src 也编码一种酪氨酸激酶,但活性还不到 v-src 激酶的 10%。

对于 c-src 和 v-src 的起源,人们已提出了两种可能性:

1. 在漫长的进化过程中,逆转录病毒反复地感染这些生物,从而使它们体内也出现了 c-src。这就是说, c-src 是 v-src 的衍生物。

2. 病毒在一个事件中以某种方式从细胞中获得 v-src,并由此产生了致癌能力。这就是说, v-src 是 c-src 的衍生物。

事实上,RSV 的 src 缺失突变体可以正常地增殖,这说明第二种解释可能是正确的,对 c-src 和 v-src 结构的研究也证明了这一点。逆转录病毒基因没有内含子。对细胞和病毒的 DNA 的序列分析表明: c-src 序列总是不连续的,而 v-src 是连续的。失去内含子的例子已屡见不鲜,而获得内含子的情况则还未见报道。因此,人们断定, v-src 是在很久以前的某一时候从含有 c-src 的细胞里获得的转变体。有些实验的结果使人们推测:成熟的 c-src mRNA 的 DNA 拷贝是这个过程的一个中间物。c-src 存在于多种脊椎动物中,这说明它在这些动物中是一个必需基因,多半在正常生物体的发育中起着调控作用。

其他逆转录病毒的致癌基因在细胞中也有其对应体。这些细胞基因都被间隔序列 (intervening sequences) 中断,因而具有典型的细胞基因结构。逆转录病毒从正常的细胞基因(而且还可能是必需基因)中“挑选”出致癌基因似乎是一种普遍现象。我们把这些



对应于致癌基因的细胞基因称为原癌基因 (protooncogenes)。

膀胱癌的致癌基因与原癌基因只差一个碱基。这是一种普遍现象，还是致癌基因可能与细胞中的对应体(即原癌基因)完全相同呢？搞清楚这个问题是很重要的。实际上这两种情况都已经观察到了。当两者序列确实不同时，人们可以认为原癌基因本身并不会引起转化，只有通过突变作用才可致癌。但是当两者序列完全相同时，我们就必须回答这样一个问题：为什么这个基因在正常细胞中无害，而在癌细胞中则变得如此危险呢？人们为此提出了许多假设，但其中只有一部分假设受到实验结果的支持。这些假设为：

1. 一个单个拷贝的原癌基因是无害的，但当拷贝数增加时(如整合的病毒 DNA 可提供多个拷贝)，该基因的产物也会增多，从而使细胞的生长和分裂失去控制。这一假设不无道理。而且我们常常可见到几个拷贝的病毒 DNA 整合到宿主 DNA 中，这就使得这个假设更显得合理了。

2. 和正常细胞相比，病毒的致癌基因可以产生更多的蛋白质产物。对此有两种解释：或是因为病毒基因的调控方式与细胞基因不同；或是因为病毒中的这个基因与细胞中的该基因两者所处的位置不同。

在本章的后一部分，我们还会讨论到这两种假设。

## 致癌物和致癌基因

人们曾一度认为：所有的肿瘤都是由于存在于正常细胞中的病毒基因受到化学致癌物 (chemical carcinogens) 和辐射的激活而引起的。支持这一观点的证据为：许多能诱发转化的试剂也能使正常动物的组织培养细胞中出现逆转录病毒颗粒。人们由此得出结论，即许多细胞都含有潜在的逆转录病毒，它们可被致癌物激活。有些癌症是可能由此造成的。多年以来，虽然人们一直在努力地证明这一假设，却始终未能如愿。而且，在许多种癌症，特别是人体癌症中，或是未发现病毒产生，或是根本就没有病毒以任何方式参与。在许多实验中，用致癌物处理正常细胞，虽然转化发生了，但是并没有看到病毒颗粒。

许多致癌物为诱变剂(第十一章)。这说明致癌物之所以能致癌，是因为它能使原癌基因产生突变，从而转化为致癌基因。为了验证这一假设，人们最近做了如下这样一个实验：用致癌物亚硝基甲替尿 (nitrosomethylurea) 处理大鼠后，在许多大鼠中形成了乳腺癌。将此种肿瘤切下来，再把从癌细胞中分离出的 DNA 切割为片段，然后将这些 DNA 片段加到小鼠的 3T3 的组织培养细胞中(在证明人膀胱癌基因的存在时，人们用了同样的实验)以测定其转化能力，结果出现了转化细胞。对 DNA 片段进行分析，发现转化的 DNA 含有大鼠乳腺癌的致癌基因。在正常的大鼠组织中，人们寻找并发现了该基因的对应体。通过测定乳腺癌的致癌基因和细胞的对应体基因序列，我们得到两个突出的结果：(1) 乳腺原癌基因和膀胱原癌基因具有相同的序列；(2) 乳腺的致癌基因与对应的细胞原癌基因只差一个碱基，而且碱基改变的所在位置与人膀胱的致癌基因所改变的碱基的位置相同。因此可以断定：(1) 人膀胱癌和大鼠乳腺癌是由同一基因突变造成的；(2) 致癌物使大鼠正常基因出现了突变，而引起乳腺癌。在此实验中，致癌物使正常基因突变为致癌基因。继这个实验之后，人们又观察到了许许多多通过这样的致癌物形成致癌基因的基因突变实例。

# 动物细胞中的致癌基因和病毒的致癌基因之间的关系

如果致癌基因是正常基因的变型(或不正常的表达形式),而这些正常基因又是参与细胞的生长调控的,那么,致癌基因的种类就应该很有限。因此如果我们对大量的肿瘤病毒和肿瘤细胞进行致癌基因的筛选检测的话,那么可以期望同样的一个基因会出现许多次,即许多的致癌基因会完全相同。在前一部分,我们已看到了这样的例子:人膀胱癌的致癌基因和大鼠乳腺癌的致癌基因是完全相同的。而且,因为这两种致癌基因是分别从人和大鼠的肿瘤中得到的,测定却是在小鼠 3T3 细胞中进行的,结果同样地可引起转化。因此,认为致癌基因转化细胞的能力不受物种的限制(至少在哺乳类是这样)。这不仅说明各种哺乳动物的生长调控可能有许多共同之处,而且当发现在正常细胞中也有 c-src 时,人们认为病毒和正常细胞还有可能具有相同的致癌基因和原癌基因。

致癌基因的种类是否确实很有限呢?为了回答这一问题,人们分离大量的逆转录病毒,并从人体组织中分离出致癌基因;同时,也从正常组织中分离出对应的原癌基因;再通过序列测定,发现各种致癌基因的确非常相似。最初检测的 Kirsten 肉瘤病毒就是一例,它能在大鼠中产生肉瘤。在 Harvey 肉瘤病毒中,人们发现一种类似的致癌基因,称为 ras (大鼠肉瘤 rat sarcoma)。ras 基因与前文所述的人膀胱癌致癌基因只差一个碱基。在人肺和大肠的肿瘤中也发现有 ras 基因。后来在人的神经胚质瘤(neuroblastoma)和纤

表 22-1 某些致癌基因和它们的来源

致癌基因	来 源*
abl	小鼠白血病,人白血病细胞
B-lym	小鸡及人的淋巴细胞
erbA	小鸡白血病细胞
erbB	小鸡白血病细胞
ets	小鸡白血病细胞
fes	猫肉瘤
fgr	猫肉瘤
fms	猫肉瘤
fos	小鼠肉瘤
fps	小鸡肉瘤
Ha-ras	大鼠肉瘤,人及大鼠的肿瘤细胞
Ki-ras	大鼠肉瘤,人的肿瘤,肉瘤及白血病细胞
mil	小鸡肉瘤
mos	小鼠肉瘤,小鼠白血病细胞
myb	小鸡白血病,人白血病细胞
myc	小鸡白血病,人淋巴瘤细胞
N-ras	人白血病及肿瘤细胞
raf	小鼠肉瘤
rel	火鸡白血病
ros	小鸡肉瘤
sis	猴肉瘤
ski	小鸡肉瘤
src	小鸡肉瘤
yes	小鸡肉瘤

\* 来源是各种逆转录病毒,除非注明是某种细胞的。



维肉瘤(fibrosarcoma)中又发现有一种类似序列,但变化的碱基不同。人们把这个序列称为 N-ras。通过大量地研究病毒和肿瘤(人体大肠、肺、胰脏、皮肤、乳腺、脑和白细胞的肿瘤,以及许多种大鼠和小鼠的肿瘤),人们发现了 20 多种不同的致癌基因。我们把这些基因列在表 22-1 中。它们的名字很特别,都是病毒或来源组织的缩写或首写字母的缩写。人们对这些基因的产物进行了研究,发现有几个为酪氨酸激酶(如 src),有两个为苏氨酸激酶,少数几个为 DNA 结合蛋白,还有一些与生长因子有共同之处(在本章后面讨论)。

## 致癌基因和致癌的多重步骤

到现在为止,从我们所介绍的实验来看,似乎只需要存在着有活性的致癌基因,就能使细胞发生癌变。实际情况并非如此。有许多证据表明,引起癌变至少包括两步:初始和启动(initiation and promotion)。尽管两者的界限还不甚清楚,而且关于它们的分子基础虽发展甚快却并无定论,但人们普遍认为,初始是获得突变的过程,是引起癌变的早期步骤,而启动则是晚期步骤,在这一阶段,突变得以表达。在诱发癌变的过程中,生长调节的失控是必不可少的一步。这可能是启动所造成的,但也许还涉及其他步骤。请注意,测定致癌基因转化潜力的所有试验都是使用小鼠的一种确立细胞株,即 NIH3T3 细胞。这一点是我们了解致癌基因如何引起转化的机理的关键。这个细胞系已被维持数千代了,选用它做试验,是因为它多年来一直被用来研究病毒转化和化学致癌。在此我们有一点想法提请读者注意,3T3 细胞很可能已经经历了癌变的一步或几步早期步骤。这不是凭空想象,因为 3T3 细胞的确不正常,它们在培养基中可无限地生长。

癌变是否包括多重步骤呢?为了回答这个问题,我们用一个已知的致癌基因去转化那些还没有开始无限生长的细胞,看看是否能够转化成功。但很不幸,得到很不一致的研究结果,而且这些结果还取决于所用的鉴定转化状态的标准。例如,在一项研究中,人们用刚刚分离出来的正常小鼠皮肤细胞(还不能无限永久地生长)为实验材料,所定的转化标准为接触抑制,可在悬浮状态下生长。在此条件下,在培养基中 ras 基因不能转化该细胞。但是如果先用 X 射线或化学致癌物处理皮肤的组织培养细胞,从中分离出几个能永久生长的细胞株,这时再用 ras DNA 去转染,就可形成转化细胞。由此得到启发:ras 基因,或更准确地说,ras 基因的产物决定了致癌过程的一个较晚期的步骤。但是,在另一些实验中,把能在低浓度血清中生长的能力作为转化标准,用 ras 去转染原初细胞,那么 ras 就可以转化原初细胞。目前阶段,对于 v-ras,我们只能说它可以诱发病变过程的一步或几步,对此我们实在是所知甚少\*。

有些逆转录病毒的致癌基因可在动物中致癌,但用小鼠 3T3 细胞测试时,却不展现为致癌基因。鸡白血病病毒的称为 myc 的致癌基因就是如此。myc 不能转化 3T3 细胞,但对其他细胞系却有转化作用。然而,用 myc DNA 和 ras DNA 依次地处理小鼠的原初皮肤细胞时,就可发生转化。在这个实验中,myc 代替了前面介绍过的 X 射线或化学致癌物的作用。这说明癌变至少包括两步,而且所需步骤的多少随细胞类型的不同而不同。在这方面,我们的知识少得可怜,对于这些现象目前我们理解得还远远不够。

---

\* 致癌基因的研究是那么重要,其研究进展迅猛惊人,人们也正提出各种有关的理论。

请注意,能转化 NIH 3T3 细胞的致癌基因并非都具有 ras 的碱基序列,这一点很重要。因为许多种致癌基因都转化 3T3 细胞,显然细胞中许多不同事件及过程都能导致这个细胞系转变为癌细胞。也就是说,癌变可以通过许多不同的途径而实现。

## 致癌基因的活化

我们曾经指出,正常细胞中有致癌基因的对应体,而且它们与致癌基因之间只差一个碱基对。在这种情况下,致癌基因是由正常基因突变而来的。尽管对致癌基因的产物的生物学功能还不甚了解,我们还是应该对其致癌过程作一些合理的推测:

1. 碱基的变化可增强或降低致癌基因蛋白的活性从而使基因活化。例如,如果该蛋白质是起调节作用的,它可以开动某生长蛋白质的合成,当该致癌基因的产物活性升高时,就会导致此生长蛋白质的过量合成。与此相反,致癌基因的产物也可能关闭生长抑制因子的合成,在此情况下,当该致癌基因的产物失活时,其抑制活性也消除了,结果也会导致生长蛋白质的过量合成。

2. 致癌基因的产物可能是一种促进生长的分子,它的活性是由其他的(未知)蛋白质调控的。该致癌基因的蛋白质产物发生改变(由碱基对变化造成),使它失去对调控蛋白作出反应的能力,不再受控于原来的调节子。在这种情况下,调节子不起作用,这个致癌基因的产物促进生长的活性就会变得过量。

3. 致癌基因的产物可能不稳定,它的活性由合成与分解这个蛋白质的相对速度来调控。当它的结构改变时,可能会阻止降解或降低降解速度,但不影响其生化活性。在这种情况下,细胞中该蛋白质的总活性将上升。

当致癌基因与其细胞中的对应体完全相同时,上述假设机理对于表 22-1 中的各个致癌基因的作用就不一定适用了。在细胞生物学的研究中,人们已很清楚地认识到,有些生物学功能不是简单地取决于特殊基因产物的有无,而是取决于该产物的浓度。因此,基因产物浓度发生变化,可使原癌基因转变为有活性的致癌基因,并诱导出癌变状态。现在,我们再来考虑这种浓度改变发生的机理,并提出有关的证据。

细胞中某基因产物的浓度是由其合成速度决定的(当然,也由降解速度决定,但在此,我们不考虑这种情况)。大多数蛋白质的合成速度是由其 mRNA 的产生速度决定的,而 mRNA 的合成又受到 DNA 结构的种种特性的影响。例如,启动子的强度和调节区。很明显,当启动子的强度和调节区两者中有一个或两者都改变时,与其相邻的基因的活性也会发生相应的改变。人们也的确观察到这种现象:有些致癌基因和正常对应体相比,只在启动子部分有碱基变化。推测这种变化可能就改变了 RNA 合成的速度。

另一种重要的机理指出,若把原癌基因移位,如把它移到另一个启动子边上,或者把它和原来相邻的调节序列分开,也可改变此原癌基因的表达。这可能就是为什么原癌基因被放到病毒中就变成了致癌基因,也就是说,正常细胞中与原癌基因相邻的启动子和病毒中的启动子相比,其强度相差很多。或者,病毒中有增强子(这种情况在 DNA 病毒中出现的可能性比在逆转录病毒中更大些)。因此,当病毒 DNA 插到被感染细胞的染色体中时,病毒的致癌基因的表达速度要比原癌基因高得多,因为原癌基因还处于它的正常位置上(图 22-5)。



许多实验表明,原癌基因位移 (relocation) 后可变为致癌基因。例如,用遗传工程的方法,将正常的小鼠原癌基因 *c-ras* (另一个为 *c-mos*) 接到病毒的调节序列上,但并不改变其碱基序列(就像 *v-ras* 一样),这样得到的遗传工程的 DNA 就能转化小鼠 3T3 细胞。在另一个实验中,人们分离出了含有原癌基因的 DNA 片段,把末端序列除去(推测这些末端序列中可能有邻近于基因的调节序列的部分),这样得到的原癌基因也有能力去转化小鼠 3T3 细胞。

癌细胞的标志之一是染色体发生改变,其中最常见的是易位 (translocation)。在易位过程中,两个染色体进行了明显的部分交换。例如,在 Burkitt 淋巴瘤中,90% 的病人第 8 和第 14 号染色体进行了交换,5% 的病人第 8 和第 2 染色体进行了交换,在另外 5% 的病人中,第 8 和第 22 染色体进行了交换。请注意,在所有情况下,都移动了第 8 染色体上的片段。原癌基因 *c-myc* 就位于

第 8 染色体上,在所有的易位过程中,*c-myc* 都移到某个编码抗体的基因的旁边,这就使原癌基因受控于一个新的调节基因之下,这个基因原来是调节抗体合成的,在这种情况下,原癌基因产物被大量地合成(但这是否对说明 Burkitt 淋巴瘤的病因有很大意义,并不清楚)。人们已发现许多原癌基因易位的例子。例如,慢性骨髓性白血病人 (chronic myelogenous leukemia) 的第 22 和第 9 染色体之间发生了易位,原癌基因 *abl* (也是 Abelson 小鼠白血病毒病毒的致癌基因)从它所在的正常位置第 9 染色体上移到了第 22 染色体上。*abl* 在这个新位置上,使 *c-abl* mRNA 的合成速度比原来增长了 8 倍。

我们不应该产生这样一种错觉,即只是增加某原癌基因的 mRNA 数量就足以致癌(我们曾讲到癌变包括两步)。实际上,在好几个实验中,人们用遗传工程的方法使原癌基因 mRNA 的合成增强,但这并没能引起癌变。这说明,癌变发生的条件不仅于此,但具体还需要什么尚有待研究。

## 致癌基因蛋白质

致癌基因和原癌基因都编码蛋白质,这些蛋白质必然与转化有关。在这一部分,我们将研究几类致癌基因蛋白质,在可能的情况下,也讲一讲它们的作用机理,即它们是如何工作的。

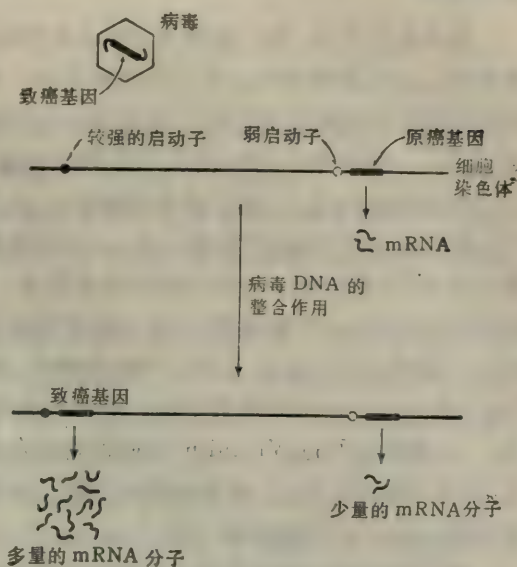


图 22-5 使病毒的致癌基因产生过量产物的一种方式。当病毒基因的编码序列与其细胞中的对应体的序列相同时,病毒基因被整合到一个比细胞的原癌基因启动子更强些的邻近的启动子旁,可以使病毒的致癌基因达到以上的结果。另一种达到以上结果的方式,就是强的启动子可能本来就存在于病毒 DNA 之中。

## 蛋白激酶

在本章的开始,我们提到几种致癌基因编码的蛋白激酶,它们可以使酪氨酸或苏氨酸磷酸化。这些激酶为什么可以引起转化呢?为了搞清这个问题,对那些含有磷酸化酪氨酸或磷酸化苏氨酸的蛋白质,需要看看它们到底有什么功能。虽然这类研究并未达到预期目标——搞清转化过程,但是我们确实知道了为什么伴随着转化作用会出现种种变化。在被劳氏肉瘤病毒(RSV)感染的细胞中,人们发现了第一个致癌基因蛋白,定名为P36。P36位于细胞膜内侧(细胞质一侧)。在某些类型的细胞中,P36以某种方式与微丝的泊繫定位有关(微丝是一些棒状的细胞质纤维,由肌动蛋白构成,而肌动蛋白则与维持细胞的形状有关)。在癌细胞中,微丝的排列是杂乱无章的,对观察到的这一情况的意义目前还不清楚。Vinculin是另一个含有磷酸化酪氨酸的蛋白质,它处于细胞表面的小斑块(粘合斑)中,这些小斑有利于细胞与其他表面或其他细胞相粘合,并可作为微丝束的附着点。但磷酸化了的Vinculin则使得它组合微丝的能力下降,这种下降变化很可能是对于“癌变”作出的反应。癌变的特点表现为癌细胞之间的粘和力低于正常细胞,而且癌细胞的形状比正常细胞更圆一些。

磷脂酰肌醇是细胞膜的一种成分,它能被src激酶磷酸化,从而被切为两部分:二酯酰甘油和磷酸肌醇。二酯酰甘油可以激活一种丝氨酸激酶,从而使许多蛋白质发生变化。而磷酸肌醇则可使结合在膜上的 $\text{Ca}^{2+}$ 释放出来。我们还不清楚 $\text{Ca}^{2+}$ 的释放在此处究竟有什么意义,但许多癌细胞中的 $\text{Ca}^{2+}$ 代谢都是不正常的。磷脂酰肌醇的两个分解产物所造成的这两类变化可能都会改变对于细胞分裂的调控作用以及对于转化作用中的其他过程。

## 生长因子

各种物质都可影响特殊类型的细胞生长速度,它们中的一些被称为生长因子(growth factor)。最近,人们发现致癌基因产物和生长因子之间有联系。鸟类红细胞溶血症(avian erythroblastosis)病毒可导致红细胞细胞瘤。这个病毒有一种叫做v-erbB的致癌基因。这个基因很大,其中一部分编码酪氨酸蛋白激酶。然而,令人感兴趣的是该蛋白质的另一部分,因为它和细胞表面的上皮生长因子(epidermal growth factor, EGF)的受体很相似。这个因子的正常受体是由v-erbB的细胞对应体——c-erbB基因编码的。这个正常受体是一个很大的蛋白质,能够横穿整个细胞膜。v-erbB只编码该受体的一部分(位于膜中的和细胞内的那两部分),病毒只获得了c-erbB基因的一部分。在有些细胞调控体系中,这个受体在传递信息中起着至关重要的作用。v-erbB所产生的变异受体很有可能扰乱这种细胞调控体系。有趣的是,当EGF结合到这种变异的受体上时,酪氨酸激酶就被活化。读者肯定会感到在所观察的这些现象中,里面一定有某些意义重大的事件,但是现在还没有揭露出来。

致癌基因v-sis存在于猿猴肉瘤病毒中,它所编码的蛋白质与原癌基因c-sis编码的血小板衍生生长因子(platelet-derived growth factor, PDGF)相似。被猿猴肉瘤病毒感染细胞不断地产生变异的PDGF,它又可与细胞表面受体结合;但是并不是所有的细胞都能产生PDGF受体。因此,很可能这种转化作用只能发生在那些已经产生出受体的



细胞受到感染时。这样可能引起自发促进的细胞分裂。这种假说,目前我们还没有实验证据。

## **ras 蛋白**

生物界的一种基本主题是在所有的生物中虽然原核和真核生物略有差别,但是基本的生命过程通常都以极其相似的方式进行。例如,所有的生物都用 DNA 和 RNA 聚合酶合成 DNA 和 RNA,糖酵解是葡萄糖代谢的普遍途径,原核和真核生物的蛋白质合成只在一些细节上有所不同等。因为细胞分裂是所有细胞的共性,因此人们猜想,各种细胞在细胞分裂的机理和调控上也必然有共同之处。

人们对细菌和动物的细胞分裂都进行了研究,但前者进展较快。在一定的时间和培养条件下,细菌分裂的频度比动物细胞高约 50 倍,而且细菌只有一个染色体,其分裂机理要比动物细胞有丝分裂的复杂体系简单得多。细菌和动物细胞在增殖上有一个根本的差异,即细菌的细胞分裂是连续的,而组构化了的动物细胞则不是这样。酵母细胞以及其他单细胞真核生物则介于两者之间;它们像细菌一样,也是自由生活的生物体,但它们又总是真核生物,具有多条染色体。因为原核和真核生物在数千万年前进入了不同的进化途径,所以人们有希望在酵母中发现某些与动物细胞相对应的细胞分裂调控机理。如果事实果真如此,那么人们就可以通过实验方式去对细胞分裂的调控进行了,因为用酵母作为实验材料要比用动物细胞方便得多。最近,在酵母和果蝇中,人们检测到 src 基因和 ras 基因类似物。酵母中有两个 ras 基因,当两者都因突变而失活时,酵母就会停止生长;如果任意一个未失活,酵母仍然正常生长。但是,如果两个 ras 基因都失活的酵母再度获得含有人的 c-ras 基因的 DNA 片段时,酵母生长又得以照常进行。这是 80 年代最令人兴奋和发现之一,这个发现给人们以信心去设想: c-ras 基因在动物细胞中起着调节细胞生长和细胞分裂的作用。

更为有趣的是,只有原癌基因才能校正酵母中的缺陷,而那些变异形式,例如各种逆转录病毒中一些不同形式的、单个碱基被改变的基因,则无此作用,它们都不能取代一个有缺陷的酵母基因。由于这一现象发现还不久(1985 年),所以人们对正常的 ras 基因的功能还了解得不太清楚。此外,人们发现正常的 ras 蛋白可激活产生 cAMP 的腺苷酸环化酶,这一现象是很令人深思的。

## **DNA 病毒中的转化蛋白质**

DNA 肿瘤病毒、腺病毒、多瘤病毒以及 SV40 病毒都可产生转化蛋白质。

被腺病毒转化的细胞中只含有一部分腺病毒 DNA。我们前面曾说过,腺病毒 DNA 中有一个特定的区域出现在被转化的细胞中。用限制性内切酶切下含有上述特定区域的片段,其中包括 E1A 基因和 E1B 基因,用它去转染细胞可诱导出转化作用。人们认为这两个基因是致癌基因。然而,这些基因所编码的产物是对于 DNA 复制必不可少的早期蛋白质,因此,它们在根本上不同于逆转录病毒所产生的非必需的致癌蛋白质,因为后者对其本身的生存来说是可有可无的。

多瘤病毒的 a 基因及 SV40 病毒的 A 基因都编码 T 抗原(见第二十一章),在受纳宿主(permissive host)中的 DNA 复制以及在非受纳宿主(nonpermissive host)中的转

化都离不开 T 抗原。但 DNA 复制和转化的关系实际上比人们早期预想的要复杂得多。这有许多原因,其中之一就是这些遗传位点都编码着好几个蛋白质(如图 21-24 所示的 SV40)。其编码区域是互相重叠的。SV40 的早期区域编码两个蛋白质: t 抗原和 T 抗原,多瘤病毒编码三个蛋白质: t 抗原,中 t 抗原和 T 抗原。人们在研究 SV40 的 t 和 T 区中的 A 基因突变时发现,在有些细胞系中,转化作用只需要 T 抗原(在某些细胞系中,转化还需要 t 抗原,但原因不明)。T 抗原是具有多种功能的蛋白质,其分子的不同区域有着不同的功能。人们曾分离到一种 T 区域有缺陷的突变体:它在非允许温度下不能进行 DNA 复制,但却可使细胞转化。因此,对于 SV40 病毒来说,转化不一定需要 DNA 复制,而多瘤病毒的情况要复杂得多。

在生物界中可能有好几种类型的致癌基因,如 myc 型和 ras 型。因此,人们对于 DNA 病毒所引起的转化的认识与理解也有了一点进步。多瘤病毒的 T 区和腺病毒 E1A 基因似乎与 myc 很相似,而多瘤病毒的中 t 区和腺病毒的 E1B 基因则与 ras 类似。

非常有意义的是,有些在转化中具有完全相同功能的 DNA 病毒基因,并不是都具有相同的碱基序列。这再次证明,不同的基因可以有同样的功能。

人们一直在那些被 DNA 病毒转化的细胞中寻找病毒诱导的蛋白激酶,但是至今还没有得到确定的结果。

## 参 考 文 献

- Axel, R., T. Maniatis, and C. F. Fox. 1979. *Eukaryotic Gene Regulation*. Academic Press.
- Bostock, C. 1980. "A function for satellite DNA?" *Trends Biochem. Sci.*, 5, 117—119.
- Britten, R. J., and D. Kohn. 1968. "Repeated segments of DNA." *Scienc. Amer.*, April, pp. 24—31.
- Brown, D. 1981. "Gene expression in eukaryotes." *Science*, 211, 667—674.
- Davidson, E. H., and R. J. Britten. 1979. "Regulation of gene expression: possible role of repetitive sequences." *Science*, 204, 1052—1059.
- Gough, N. 1981. "The rearrangement of immunoglobulin genes." *Trends Biochem. Sci.*, 6, 203—208.
- Gurdon, J. B. 1974. *The Control of Gene Expression in Animal Development*. Harvard.
- Herskowitz, I., and Y. Oshima. 1982. "Control of cell type in *Saccharomyces cerevisiae*: mating type and mating-type conversion." In J. Strathern, E. Jones, and J. Broach eds. *The Molecular Biology of the Yeast Saccharomyces*. Cold Spring Harbor Laboratory.
- Hood, L. E., I. L. Weissman, and W. B. Wood. 1978. *Immunology*. Benjamin.
- Kedes, L. H. 1979. "Histone genes and histone messenger." *Ann. Rev. Biochem.*, 48, 837—870.
- Kolata, G. 1981. "Gene regulation through chromosome structure." *Science* 214, 775—776.
- Leighton, T., and W. F. Loomis, (eds.) 1982. *The Molecular Genetics of Development*. Academic Press.
- Lewin, B. 1981. *Gene Expression*. 2. *Eukaryotes*. Wiley.
- Long, E. O., and I. B. Dawid. 1980. "Repeated genes in eukaryotes." *Ann. Rev. Biochem.*, 49, 727—766.
- Marx, J. L. 1981. "Antibodies: getting their genes together." *Science*. 212. 1015—1017.
- Ochoa, S., and C. de Haro. 1979. "Regulation of protein synthesis in eukaryotes." *Ann. Rev. Biochem.*, 48, 549—580.
- O'Malley, B. W., and W. T. Schroeder. 1976. "The receptors of steroid hormones." *Scienc. Amer.*, February, pp. 32—43.
- O'Malley, B. W., H. W. Towle, and R. J. Schwartz. 1977. "Regulation of gene expression in eukaryotes." *Ann. Rev. Biochem.*, 46, 239—275.
- Revel, M., and Y. Goner. 1978. "Posttranscriptional and translational controls of gene expression in eukaryotes." *Ann. Rev. Biochem.*, 47, 1079—1126.
- Safer, B., and W. F. Anderson. 1978. "The molecular mechanism of hemin synthesis and its regulation in reticulocytes." *CRC Rev. in Biochem.*, 6, 261—290.
- Sakano, H., H. Hüppi, G. Heinrich, and S. Tonegawa. 1979. "Sequences in the somatic recombination of immunoglobulin light chain genes." *Nature*. 280, 288—294.



- Seidman, J. G., and P. Leder. 1978. "The arrangement and rearrangement of antibody genes." *Nature*, 276, 790—795.
- Tata, J. R. 1976. "The expression of the vitellogenin gene," *Cell*, 9, 1—14.
- Wahli, W., I. B. Dawid, G. U. Ryffel, and R. Weber. 1981. "Vitellogenesis and the vitellogenin gene family." *Science*, 212, 298—304.
- Weisbrod, S., and H. Weintraub. 1979. "Isolation of a subclass of nuclear proteins responsible for conferring a DNase I-sensitive structure in globin chromatin." *Proc. Nat. Acad. Sci.*, 76, 630—634.

## 第二十三章 真核生物基因及基因产物活性的调节

原核细胞与真核细胞的调控有很大差异:原核细胞一般是自由生活的单细胞有机体,只要有合适的环境条件和充足的养料供应,原核细胞能无限地生长和分裂。因此,它们的调控系统的作用是在一个特定环境中为细胞提供最大的生长速度,除非生长受到伤害。这个最佳生长是通过下面的机理获得的,即通过将细胞的合成和酶的活性控制到最大生长速度所要求的点上,也就是去除合成中的废物,并使细胞能尽快地适应变化着的环境条件。因为原核细胞 DNA 无核膜包裹, DNA 可以很快地接收到细胞质中的信号。这样,其蛋白质合成开关经常是通过控制转录的起始来调节的。这方面在第十四章中介绍的几个操纵子已可看到。

除了酵母、藻类和原生动物等单细胞真核生物以外,其他各种真核细胞对基因表达的调控有不同的要求。在一个发育着的生物(如在胚胎)中,一个细胞不仅要生长、分裂,产生很多子代细胞,而且这些子代细胞在形态上和生化方面发生很大变化,并为维持这种变化状态,它们还必须发生分化。然而,这些细胞的生长和分裂却比细菌简单,因为它们生长所处的介质的成分和浓度比较稳定,这些介质为血液、淋巴液或其他体液等。对于海生动物,介质则是海水。在一个成年有机体中,大多数类型的细胞都停止生长和分裂,每一个细胞只需维持着它们自己。还有很多这样的例子。重要的一点是由于典型的真核细胞和原核细胞的要求不同,所以它们的调控机理基本上是不同的。由此,人们可能认为自由生长的单细胞真核生物的调节机理类似于原核生物,但实际上并不是这样。在远古时期,核区被膜包围是进化中的一个很大的进步。这一进步是由基因的结构及调控发生重大改变而完成的,或者更可能是在基因的结构及调控发生重大改变之前就已完成。

这一章主要研究形成一定结构组织的真核细胞,大部分讯息来自于对哺乳动物、两栖类动物(*Xenopus*, 爪蟾类的蟾蜍)、昆虫(*Drosophila*, 果蝇)、鸟类(通常是鸡)和棘皮动物(sea urchin, 海胆)的 DNA 进行的详细研究;此外,对于啤酒酵母(*Saccharomyces*)的基因表达的研究,也得出一些不寻常的讯息。

多细胞真核细胞基因表达调节的研究曾因两方面问题受到阻碍,即无法得到调控的突变株,不能对真核基因进行分离与操作,而且缺乏体外的 RNA 聚合酶 II 转录系统。近来,重组 DNA 技术的兴起改变了这一切。这种技术可以将 DNA 片段进行克隆(见第二十章),克隆了的 DNA 片段可以用来研究基因结构,并可作为探测 mRNA 的探针。已获得的资料大大加快了研究的进展。虽然现在对于一些事实的概貌还没有弄得很清楚,但是已掌握了大量的讯息材料。因此这一章将对于采用各种不同的细胞所观察到的、多种多样的调控机理作概略性的介绍。

### 原核细胞和真核细胞基因结构的重要差异

原核细胞与真核细胞在基因转录、翻译以及 DNA 的空间结构方面有很大差异。这



些差异中的一部分已在前面的章节中讨论过，另一部分则将在下面的章节中讨论。其中七点差别将在讨论真核细胞基因表达的调节中重复地出现，这些差别如下：

1. 在真核细胞中，一个已完成的 mRNA 分子只能翻译出一条单个的肽链(见第十三章)，因此在真核细胞中见不到原核细胞中所存在的那种类型的操纵子。

2. 真核细胞的 DNA 与组蛋白和大量的非组蛋白结合在一起(形成染色质)，只有一小部分 DNA 是裸露的。

3. 真核细胞中有一部分很重要的 DNA，它们是由几个碱基组成的短序列，重复几百次—几百万次形成重复序列。

高等真核 DNA 中的很大一部分是不被翻译的。

4. 真核细胞能够用有控制的方式使 DNA 片段进行重排。还具有能在需要时增加基因数量的潜力(这种潜力在原核生物中也有，但似乎很少)。

5. 在大部分真核细胞的 DNA 中存在着内含子。

6. 在原核生物中，转录作用的调节位点都很小，它们靠近而且通常就在启动子的上游，蛋白质结合到调节位点上可直接地促进或抑制 RNA 聚合酶对它的结合。在真核生物中，调节区则大得多，它们可能远离启动子达几百个碱基之多。这些调节区也能结合某些蛋白质，但是这些蛋白质离启动子太远以致不能在结合的同时对启动子起影响作用。

7. 在真核生物中，RNA 是在细胞核中合成的，但是它必须经运输穿过核膜到达细胞质后，才在细胞质中进行翻译。这样严格的空间间隔的影响在原核生物中几乎不存在。

纵观本章，我们将看到以上这些性质是如何地表现于特殊的调节模式中的。

## 基因家族

在原核细胞中，密切相关的基因往往组成操纵子并且以多顺反子 mRNA 的方式进行转录。这样，整个体系就置于一个启动子的控制之下，通过控制这个启动子就可以对整个体系进行开启和关闭。因为真核 DNA 都是单顺反子，所以，如在真核细胞中对这些不加以改变，那么上述调控方式就将是无用的。在真核细胞中，许多相关的基因可以按功能组合成一套基因，这一套基因被称作基因家族。此家族中的各个成员的活性虽然有时不很协同，但通常是协同着的。现在，基因重组技术已用于克隆基因家族；偶然，也发现过同一个家族中的所有基因都可以克隆到一个单个的质粒上面，这就很清楚地表明：这样的—个家族形成一个基因簇。但更经常的是，—簇相关基因太大以致不能整个地克隆在一个质粒中，而是家族中的两个或三个基因作为一个单位被克隆。这个不完全的发现对决定基因的顺序非常有用。例如，如果一个克隆含有 A 和 B，另一个含有 B 和 C，第三个含有 C 和 D，则基因在染色体上的顺序必然是 ABCD。克隆实验指出，还有许多基因，它们的产物相关，但是调节并不协同，这些基因也能成簇。这样，基因家族的称号就扩展了，任何功能相关的基因簇都可包括在内(例如 tRNA 基因)。

基因家族通常分为简单的多基因家族、复杂的多基因家族和发育阶段控制的复杂多基因家族。在图 23-1 中给出了每个类型的例子。

简单的多基因家族

简单的多基因家族是指在家族中有一个或几个基因以串联排列的方式重复着。这种家族的最简单的例子就是编码 5S rRNA 的一串基因,这在爪蟾中已作过详细研究。这一组 rRNA 基因形成一个巨大的阵势,其中的每一个 5S rRNA 基因都被间隔序列分开,从而形成一个基因簇。间隔序列的大小变动于 2—6 倍 5S rRNA 基因的长度之间,而且,间隔序列还含有中等重复序列。每一个 5S rRNA 基因都被单独地转录,产生出一个与别的部分之间分开的 RNA 分子,这些 RNA 分子再经过加工成为成熟的 5S rRNA。在爪蟾细胞中有好几个由几百个—几千个 5S rRNA 基因聚集而成的基因簇拷贝。

爪蟾中的 5.8S、18S 和 28S rRNA 基因组成了另一个简单的多基因家族,其中的每一个转录单位都含有这三种 rRNA 序列,每个转录单位又被间隔物分开,5.8S、18S、28S rRNA 分子在加工酶的作用下从初级转录子上剪切下来,在各种类型的二倍体细胞中,这三种 rRNA 基因序列在每个细胞中的拷贝数为几百—几千个。

rRNA 基因提供了原核生物细胞、单细胞真核生物细胞以及多细胞真核生物细胞之间有趣的差别。例如,在大肠杆菌中,5S、16S 和 23S 的 rRNA 联合组成一个转录单位,各种 rRNA 分子是从这个转录单位上剪切下来的。而在酵母中,许多 rRNA 分子形成一个单独的重复单位,每个单位由间隔序列分开。然而 5S rRNA 的转录则独立进行,与 5.8S-18S-28S 的转录单位没有关系(这种组合也许有进化意义,因为真核细胞的 5.8S rRNA 与原核细胞的 5S rRNA 的功能相对应)。最后,多细胞真核生物细胞中的 5S rRNA 基因组成独特家族,前面已经讲过。

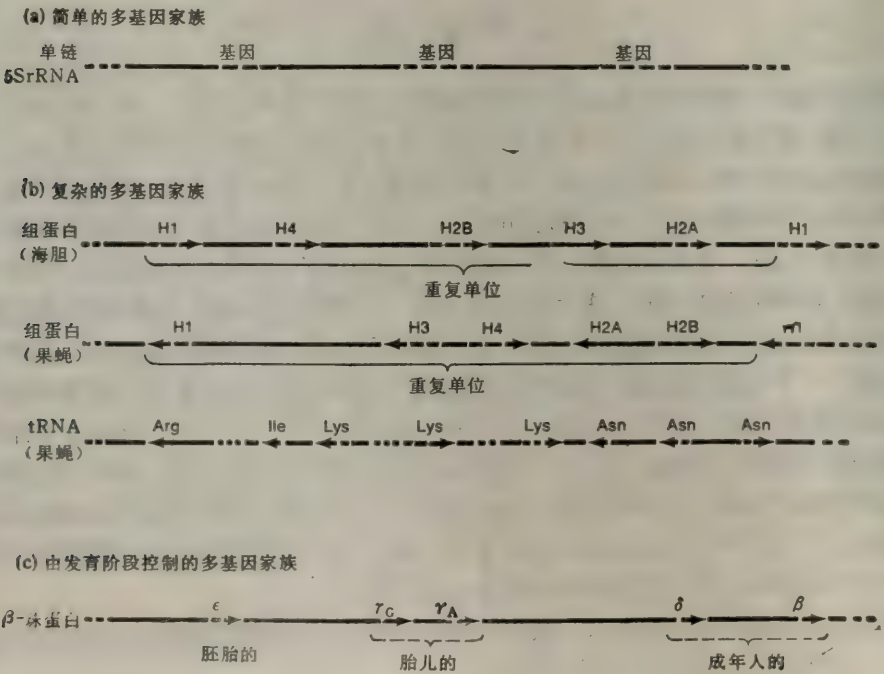


图 23-1 基因家族的 5 个例子。箭头指出转录的方向。



复杂的多基因家族

复杂的多基因家族一般由几个相关基因簇组成,每个基因簇被间隔序列隔开,独立地进行转录。在图 23-1(b) 中给出了三个例子,每个例子都有其本身的特征,所以自然界里有好几种不同类型的复杂的多基因家族。

首先,让我们看一下海胆的组蛋白基因家族。在这种动物中,由 5 个基因编码着五种组蛋白,这 5 个基因处于一个小的 DNA 片段中,而且被间隔序列分隔开来。这个 5 基因单位串联状地重复约 1000 次。5 基因单位中的每一个基因分别地被转录成单顺反子 RNA (令人吃惊的是这些 RNA 都没有内含子)。注意,各个基因都是按同一方向进行转录,也就是从 DNA 的同一条链上转录。每个基因的转录与翻译的速度都受到调节,但目前对这种调节的方式还只能作简单的讨论。首先,组蛋白在适合于染色体复制作用的时候合成出来;其次是有控制地合成出相等分子数的 H2A、H2B、H3 和 H4 以及半个分子数的 H1 (这是染色体中的分子数比例)。毫无疑问,5 基因单位具有多个拷贝可以使细胞在 DNA 复制开始时快速地合成出大量的组蛋白。最近发现,在一个特定的细胞中并不是所有的 5 基因单位的拷贝都得到转录,而是在胚胎发育的不同阶段和不同组织中,有着不同的 5 基因单位在进行转录。这暗示出在不同的发育时期,可能存在着具有不同专一性的组蛋白亚类 (后来证明这个基因家族是发育调控型的各种基因家族类别中的一个成员)。后面将对组蛋白基因的调控作更多的介绍。

检验果蝇组蛋白基因家族时,发现了它与海胆的不同之处:这 5 个基因是按两个方向进行转录的,如图 23-1 所示。而在酵母中,组蛋白基因的排列方式与海胆中的差别则更大,存在着两个相距很远的基因簇,每一个含有 H2A 基因和 H2B 基因,而其他的组蛋白基因则是分开地散布于整个染色体中。

图中所示果蝇的 tRNA 基因簇具有两个上述组蛋白簇所没有的特点,那就是在基因簇中含有同一种基因的多个拷贝(如三个 tRNA<sup>lys</sup> 的基因),而且,不是所有的 tRNA 基因都在这个基因簇之中。基因簇中的每一个基因都单独地转录,并且是按两个方向进行转录。人们还发现其他的处在基因簇中的 tRNA 基因也在 DNA 中的别处存在着,甚至存在于其他一条染色体上。tRNA 基因的成簇作用在单细胞的真核生物中并不是一个普遍性的规律。例如,酵母 tRNA 基因已知有 360 个,其中 175 个已经克隆到质粒中,只在个别情况下发现在一个重组质粒中含有两个 tRNA 基因。更进一步,八个酵母 tRNA<sup>Tyr</sup> 基因位于六条不同的染色体上。

受发育控制的复杂多基因家族

血红蛋白是一个具有两条  $\alpha$  链和两条  $\beta$  链的四聚体蛋白质。然而,在有机体发育的不同阶段,却出现有一个或几个氨基酸差别的几种不同形式的  $\alpha$  和  $\beta$  亚基。下面的时间顺序表给出了人体于怀孕后不同时期出现的不同亚基(图 23-2)。

	胚胎的 (<8 周)	胎儿的 (8—41 周)	成年人的 (出生→以后)
类 $\alpha$ -珠蛋白	$\zeta_2 \rightarrow \zeta_1$	$\alpha$	$\alpha$
类 $\beta$ -珠蛋白	$\epsilon$	$\gamma^G$ 和 $\gamma^A$	$\beta$ 和 $\delta$

类  $\alpha$ -珠蛋白在胚胎期,最先出现的  $\zeta_2$  型的类  $\alpha$  链,逐渐地被  $\zeta_1$  型所取代。在胎儿和

成人期,都出现  $\alpha$  型链。再看类  $\beta$ -珠蛋白,在胎儿期有两种类  $\beta$  型链即  $\gamma^G$  型和  $\gamma^A$  型链(在肽链的同一个位点上,一条链在此处含有甘氨酸,另一条含有丙氨酸),差不多是等分子数。在成年期,有 50 倍的  $\beta$  型及  $\delta$  型链。这些变化的总结果是:胚胎型的血红蛋白中含有两个  $\zeta_2$  型链和两个  $\sigma$  型链(即  $2\zeta_2, 2\sigma$ ); 出生之后, 98% 的血红蛋白含两条  $\alpha_2$  亚单位和两个  $\beta$  亚单位(习惯上称做血红蛋白

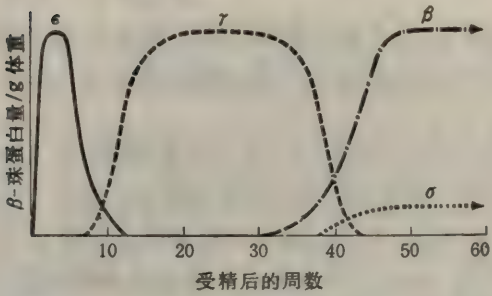


图 23-2 在人体发育的时间过程中不同的类  $\beta$ -珠蛋白链的浓度变化。

A), 2% 的血红蛋白含两个  $\alpha_2$  及两个  $\delta$  亚单位(血红蛋白  $A_2$ )。

类  $\alpha$  的和类  $\beta$  的基因都形成独立的簇,  $\beta$  簇见图 23-1(c)。每一个这样的基因簇有一个显著的特点,即基因排列的顺序就是它们在发育中表达的顺序(这一点并不适用于所有的发育控制型的基因家族)。人们现在还既不了解这些不同形式的重要意义,也不了解程序性合成方式的任何细节。

人类的  $\alpha$ -和  $\beta$ -珠蛋白基因簇已经被克隆并分析了它们的重复序列。其中某些区域的碱基序列也已确定。从上述分析中得到的最重要发现就是在单个的基因之间有重复序列。这个事实可用来支持前面提到过的重复序列具有某些调节功能的见解。

通过对不同生物珠蛋白基因簇的研究,能够了解基因簇的产生过程以及可能维持其存在的进化因素。原始鱼类、海蚕以及昆虫都有单个的珠蛋白基因,而两栖类的  $\alpha$  和  $\beta$  基因则紧密地连锁在同一条染色体上。低等哺乳动物和鸟类的  $\alpha$  和  $\beta$  基因都有多种亚型,而且两个基因簇位于不同的染色体上。由此得出假说: 原始的珠蛋白(大约产生于 8 亿

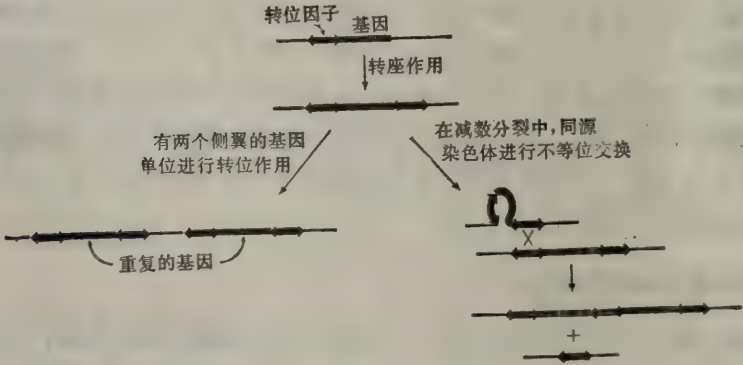


图 23-3 几种基因重复机理中的两种机理。这两种机理都从转位作用事件出发,即移动因子从染色体中的一个位点移动到另一个位点上。图中左边的途径: 带有两个移动因子序列侧翼的基因单位经过第二次转位作用,完成基因重复过程。右边的途径: 在两个转位因子之间的不等位相互交换,形成一个带有重复基因的染色体和一个缺失了一部分物质的染色体。但是,如果原初基序两侧的两个序列对于不等位交换有足够的同源性,那么右边的途径就可能在并没有转位因子的情况下发生。在右边途径中还可能有一种机理,如果在侧翼的非同源序列之间发生重组作用,那么就会发生很罕见的、非同源重组作用。



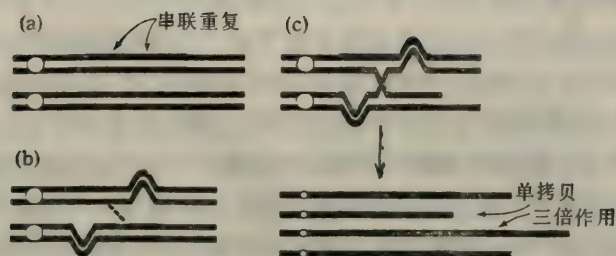


图 23-4 串联重复过程的不等位交换造成染色体片段的拷贝数增加。(a) 带有串联重复的染色体的正常配对。(b) 错配。下面的染色体的右边部分与上面的染色体的左边部分配对。虚线指出交换的潜在位点。(c) 错配复制中的交换,产生四倍的染色体,如图所示。再进一步形成一个携带着一个单拷贝复制区的染色体单体,以及另外三倍体染色体。

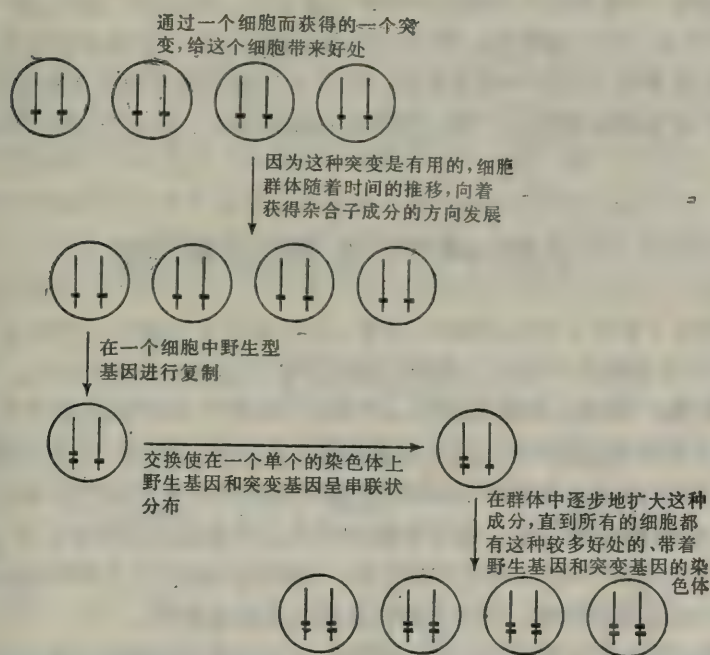


图 23-5 在整个群体中将一个有用的突变散布开来,而不丢失野生型基因。

年前)是由一个基因编码的,现代的珠蛋白基因簇则是由一个祖先基因通过一系列的基因重复、突变以及转位而形成的。有人提出了这样一个假设的顺序事件(这个假设很可能是事实):早期生物的体积很小,虽然血红蛋白由一个亚基组成,运氧能力有限,也能满足功能的需要。通过自发突变,血红蛋白基因发生了变异,从而使杂合子基因能够编码出运氧能力更强的四聚体蛋白质分子。四聚体分子对在组织中它所结合的氧的释放的比例要比单体分子大得多,因此能够进化成较大体积的动物。后来,在哺乳动物进化过程中,两条 $\beta$ 链的基因发生了突变和重复,形成了胎儿中的 $\gamma$ 型珠蛋白。胎儿血红蛋白对氧的亲合力比成人血红蛋白还要大,因而有利于胎儿的快速发育,由此有可能产生更复杂的生物。

在灵长类进化期间,还发生了进一步的突变和重复,从而产生了更多的血红蛋白的型态。

基因重复以及随之发生的重复基因扩散到整个种群的机理有特殊的遗传学意义。图 23-3 表示了理论上可给出的几种基因重复的机理。例如,基因一端的一个简单的转位因子(可运动的 DNA 序列,在第十九章中讨论过)可以转位到另一端去,并在原先的位置上留下一个完全一样的序列,从而形成一个复杂的大序列。这个复杂的大序列又可以作为一个整体再次转位到一个新的位点上。而在原先位置上仍留下一个完全一样的序列拷贝并从而产生基因重复。另一种造成基因重复的方式是通过非等位基因重复。这样可以造成同源染色体中一条上基因重复,另一条上基因缺失。进一步交换可以产生三倍重复或更多的重复,如图 23-4 所示。图 23-5 说明一个有利的突变型怎样与原先的野生型基因共存。

这里描述的事件虽然发生在很久以前,现在仍有遗迹。这一事实有力地支持了上述图式。通过对几种生物中珠蛋白基因簇中多片段的研究发现,有些无功能的序列几乎与正常的珠蛋白基因完全相同。这些序列被称为假基因(pseudogenes)。假基因通常被认为是进化的遗迹,它是曾经有功能的基因经过重复后的衍生物。推测,某一时期发生了一个突变,使这个基因的产物失活。因为这个基因至少还有一个有功能的基因副本存在,所以没有很强的选择压力来消除突变基因,以后再生成的突变就将逐渐地积累起来。假基因在基因簇中普遍地存在着。一般它经常地在转录作用、内含子切除以及翻译作用等方面有缺陷。

## 真核 mRNA 分子的多种类型

我们已讨论了原核生物不同种的 mRNA 分子的浓度因操纵子不同而不同。在“开”的状态下,最丰富与最不丰富的 mRNA 的浓度变化范围是几百个因次。这在真核生物中变化范围更大。而且,多数丰富的 mRNA 分子的合成可能是由不同于其他类型 mRNA 合成的机理来控制的。因此,有必要给三类真核 mRNA 分子下定义:低拷贝数 mRNA 分子(low-copy number mRNA),每个细胞有 1—8 个拷贝。中度拷贝数 mRNA (moderately prevalent mRNA),每个细胞中有几百个拷贝。高度拷贝数 mRNA (super-prevalent mRNA),每个细胞中可达  $10^4$  拷贝。最后这种分子由发育中的卵细胞以及只担负着合成单一蛋白质的细胞,如合成血红蛋白的红细胞合成。

对合成低拷贝数及中度拷贝数的 mRNA 分子的调节作用的了解不像对合成高度拷贝数的 mRNA 分子的调节作用那样详细。

## 真核基因表达的调节策略

从第十四章已经知道调节原核基因表达的许多机理,但是,几乎所有方式都是控制转录的起始及终止,虽然亦有翻译水平的调节,但很不普遍。第十二章中已经指出,真核 mRNA 合成比原核要复杂得多。因此,许多个控制点都有可能起到调节作用。例如初级转录子的形成, mRNA 的产生以及 mRNA 运输到细胞质中。这些控制点也还可能有几个不同水平的调节。比如,启动子的可供性取决于染色体的状态,也取决于转录因子的活性或浓度,或增强子的可供性等。由初级转录子的产生到 mRNA 合成的调节则可



能是通过控制拼接作用及多聚腺苷酰化作用而实现。在原核生物中,由于通常要迅速地对受调节基因的产物的合成进行“开”和“关”的控制,所以很少遇到对 mRNA 寿命进行调节的情况。但是在真核生物中有时需要大量的某种特殊蛋白质中而且允许有阶段较长的时间以达到所需要的量。这时,延长 mRNA 的寿命则是一种相当合理的调节方式。

正如以前已指出的,原核生物必须能对变化着的环境做出反应,因此,还必须使基因能从“开”到“关”,并能从“关”到“开”。但是在真核细胞的分化过程中,一个基因或一个基因家族通常是永久地开着或关着的。在这种低要求的生物中可用其他方式进行调节,如基因缺失,基因失活,基因扩增及基因重排等。有资料表明,许多真核生物在发育过程中通常必须使某些特殊蛋白质维持相同的量或维持恒定比例。在原核生物中这个任务由多顺反子 mRNA 来完成。真核 mRNA 是单顺反子,这就需要另外一种机理去完成。其中一种是合成出一种能够被切割成另一种蛋白质的蛋白质。前面几部分,我们将介绍这些调节策略中的每一种调节方式,并稍详细地讨论它们。

基因家族

基因改变

高度重复的 mRNA 分子即多拷贝 mRNA 的形成通常与分化阶段的某一改变有关,例如,一个成熟的红细胞能产生大量的可翻译出成熟珠蛋白的 mRNA,而它的还未变成红细胞的前体细胞却不产生珠蛋白。发生这一变化的一种方式就是基因或它的拷贝数发生永久性变化。这些变化主要有三大类,即基因的丢失、放大及重排,后面还有转录水平,转录后水平以及翻译水平的变化。这些调节方式中的每一种将在后面几部分中讨论。

基因丢失

在细胞分化时,消除某一个基因的活性的一种方式就是从细胞中除去那个基因。某些原生动物、昆虫和甲壳纲动物以非常极端的方式按这样的方法进行分化,即丢失整个染色体。在这些生物中,只有那些预定的将要产生生殖细胞的细胞中才始终维持着全套的 DNA 组分。在那些较高等的生物中并未发现有染色体丢失的调节方式。

许多原生动物含有两种类型的核——小的称为小核 (micronucleus), 大的称为大核 (macronucleus)。通常,在生殖细胞中只具有一个小核,当细胞进行分化时(很多原生动物都是这样的),小核经历着各种变化,这种分化作用的最终结果是产生拥有一个大核和一个小核的细胞。小核中含有无活性的 DNA,这只是为将来产生生殖细胞作为一个遗传信息的贮藏所。大核中的一些 DNA 分子则是所有转录采用的模板。

人们对原生动物 *Oxytricha* 的大核的形成过程进行了详细的研究,处于生殖细胞中的小核的 DNA,在分化过程中被切割成为长度为 0.5—20 碱基对的一些线性片段。小核中的大部分 DNA 在形成上述片段的过程中被彻底地降解。通过一种目前还不清楚的方式,上述产生的片段又不断重复地复制着,直到每个细胞中含有 17 000 种不同片段的 1 000 个拷贝为止。这些众多的片段进一步建成大核。在这个过程中,大核所含 DNA 是小核 DNA 的几百倍,但是丢失了一半以上的 DNA 序列。虽然最近的实验给出了很有意义的结果,但是这种有选择性的基因丢失的机理还不清楚。一个最有意义的发现是绝大多数片段的每一个末端都有相同的碱基序列。这种末端重复可能代表着一种切割位点的

结构,即像限制性内切酶的专一性的作用位点那样的特殊结构。

在高等生物中,现在还未观察到基因丢失现象。这种调节方式可能仅存在于进化程度相当低的一些生物之中,当然也不能肯定在较高等的生物中就不发生丢失,因为要在 50 000 个基因中测查出丢失一个基因则是太困难了。然而,用蛙卵所进行的下述实验结果表明,在发育过程中必需的基因是不丢失的。

在这个实验中,通过显微操作将蛙受精卵中的细胞核去除,而使细胞仍然保持完整;然后再从蝌蚪的肠细胞中取出细胞核,将这个核注射到去核的蛙受精卵中,很多经过这种手术修复过的卵能继续生长、分裂并发育成为成活的蛙。这种接受外来的已分化的肠细胞核的受精卵是否会缺乏生长发育所必需的遗传信息呢?上面的实验结果理由充分地作出以下结论:虽然换核过程有可能会丢失一些基因,但是已分化了的细胞的细胞核并不缺少任何遗传物质;若有丢失,丢失的也是一些非必需的基因。所以说分化的细胞核是全能性的 (totipotent)。

基因扩增

基因扩增是使细胞大量地产生某一种特定基因产物的一个途径。研究得最深入的例子是爪蟾卵母细胞 (oocyte of the toad *Xenopus laevis*) 发育过程中的 rRNA 基因,在发育中这个基因的数目增加了 4 000 倍。爪蟾的卵母细胞的前体,类似于所有的体细胞(非

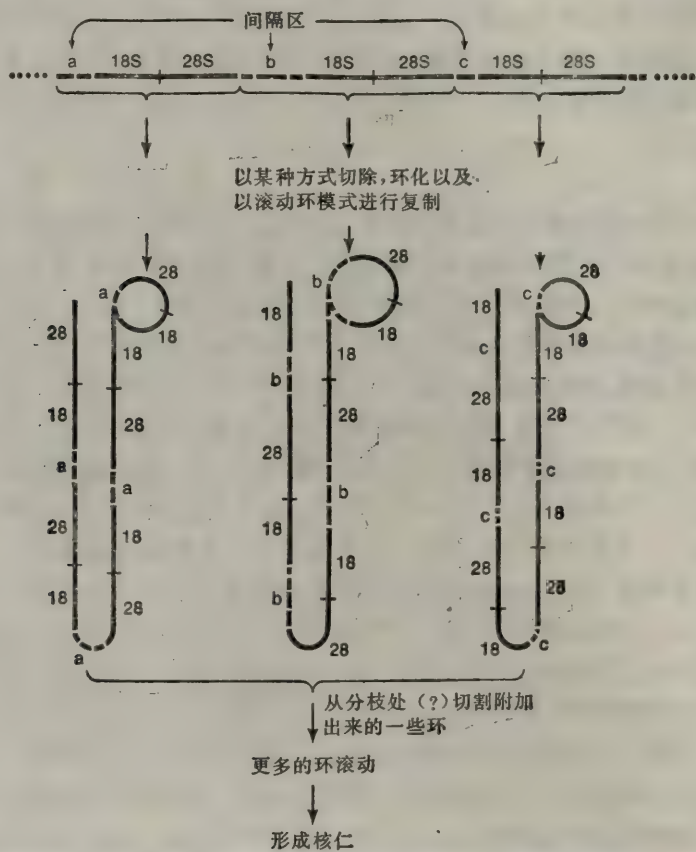


图 23-6 爪蟾 rDNA 扩增的假设机理。间隔区 a、b、c 具有不同的长度。



生殖性的细胞),含有约 500 个 rRNA 基因单位。基因扩增后每单位有  $2 \times 10^6$  个拷贝,占卵母细胞的细胞核 DNA 的 75%。这使得卵能合成出  $10^{12}$  个核糖体,满足了受精之后细胞分裂不同时期大量合成蛋白质的要求。

在基因扩增之前,rDNA 区域由串联排列着的 500 个 rDNA 单位所组成。每一个单位中含有一个编码 18S RNA 和 28S rRNA 基因。每个基因由间隔区分开(见图 23-6)。间隔区的大小不恒定,一个单位中的间隔区与下一个单位中的并不相同。整个 rDNA 位于一个单个的核仁中(人们发现大部分动物细胞中都有这样一种含有 rDNA 的核内小体)。基因扩增需用的时间,大约在三周以上,前体细胞发育成卵母细胞,rDNA 的量增加 4 000 倍,核仁的数量增加到几百个(图 23-7)。

在基因扩增期间和扩增之后,所有的 rDNA 不再由简单的连续的 DNA 片段组成,而代之出现的是大量的小环和复制滚动环。形成这些结构的精确机理还不清楚,但已有一些有关在非复制环中及复制滚动环中的 18S 及 28S 组分之间的间隔区的大小及讯息。图 23-6 是处在三个滚动环中的 rDNA 的图解。这些分子的基本特征就是每个分子含有的间隔区都是同样大小的。假如含有一些相邻 rDNA 单位的染色体 rDNA 以一个大片段的形式被切除下来并被复制,那么在一个滚动环中就应当含有不同大小的间隔区。其他几个实验的结果也表明 rDNA 单位是单个地从染色体上一一剪切下来的,如图中所示。剪切下来的分子再环化(或很可能以环的形式被剪切下来),并在通过滚动环产生出长的侧枝形式之后才进行复制的。

令人奇怪的是,在多数滚动环中,环的部分不只含有一个 rDNA 单位,而经常含 2—4 个,甚至多到 16 个。关于剪下来的一个单位的环如何转变成多单位的环的机理还不清楚,这个问题可能与下面的问题有关。一个 rDNA 单位长为  $4.3\mu\text{m}$ ,通过滚动环复制形成一个单位长度( $4.3\mu\text{m}$ )的侧枝大约需 1.5 小时。基因扩增共约进行 72 小时,在这段时间内,每个单体片段可产生 144 个拷贝,然而观察到的值却大约是 4 000 倍。所以为了合成出更多的 DNA,至少有一部分被切下来的单体必须先被复制,以产生出进一步复制所需要的模板,这可能通过几种不同途径实现。例如,一个环可能是在滚动环复制之前,通过  $\theta$  方式复制几次;不过,很少观察到  $\theta$  式分子。另一种可能性从滚动环的线性侧枝剪下几个单位的 rDNA 然后再行环化。这能够解释上面提到的存在着带有一连串串联重复的 rDNA 单位环的事实。很显然,为了说明这种现象还需要进行更多的工作。

最后,每一个复制单位组成了成熟卵母细胞中的许多核仁中的一个(图 23-7),到这个阶段,扩增就停止了。很有意思的是,在一个特定的卵细胞中不是所有的 rDNA 基因都得到扩增,而且所扩增的特定基因也随不同卵母细胞而有不同。

一旦卵母细胞成熟,多余的 rDNA 就没有用了,将逐渐降解。紧接受精之后,染色

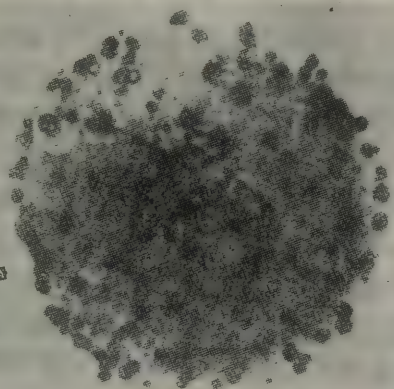


图 23-7 爪蟾卵中分离出来的核经过染色后的显微照片,显示了 rRNA 基因扩增后形成的几百个核仁(自 D. D. Brown 和 I. B. David, *Science* (1968), 160:272)。

体 DNA 开始复制,并开始进行有丝分裂,在胚胎发育过程中不断地重复这些过程。在此期间,多余的染色体 rDNA 不再复制,并继续降解,直到形成几百个细胞的时候,这种 DNA 就全都不存在了。

在许多生物卵形成过程中都观察到 rRNA 基因的扩增过程,例如,在昆虫、两栖动物、鱼以及大核形成过程中的原生动物中。

最近在果蝇中发现产生蛋白质的基因扩增作用。在卵成熟之前,卵巢颗粒细胞中产生卵壳蛋白的基因被扩增。果蝇中的一种称之为卵原细胞 (oogonium) 的前体细胞,经历 4 次细胞分裂产生 16 个细胞。其中一个是卵母细胞 (oocyte), 它注定要成为卵,其他 15 个是营养细胞 (nurse cell), 它们为卵母细胞的形成提供大量的蛋白质及其他大分子物质。营养细胞之所以能够产生大量的物质是因为在他们的形成中发生了多次特殊的 DNA 复制。这种特殊的 DNA 复制没有有丝分裂,即没有子代 DNA 分子(染色单体)的分离。也就是说,每条染色体中含有许多个相同的 DNA 分子,即多线染色体 (polytene), 如唾液腺的巨型染色体。这种染色体广泛地用于细胞遗传学的研究。扩增的第一步是染色体扩增。营养细胞与卵细胞通过胞浆桥相连,让基因产物进入正在增大的卵细胞。在卵发生的后期,营养细胞作出反应合成大量卵壳蛋白,把卵母细胞包裹在卵壳内。为了合成出所需的量,卵壳蛋白基因的扩增是通过从卵壳蛋白基因簇的复制起始点进行局部 DNA 复制,反复地起始复制作用产生网状的复制泡。如图 23-8(a) 所示。泡中多出的卵壳蛋白基因的拷贝不被切除和环化而是作为染色体的其余部分而存在。这种情况可见于 rRNA 基因,它的扩增能使细胞在短时间内产生大量的蛋白质。以后还可以看到,如果要合成大量的蛋白质而且又有较长的时间用于合成,那么基因的扩增就不很必要,因为这时可以通过延长 mRNA 的寿命而达到目的。

值得注意,这种扩增作用不能在有丝分裂中继续下去。因为完成一个复制周期以及完成子代分子的分离,最终将只能产生一个有基因簇的分子(没有“泡”)。然而在这样的体系中是不能发生有丝分裂的,因为营养细胞都是“终点”,它们在卵形成之后就脱开与卵的联系。在其他生物和其他基因中,体细胞的分化过程有时存在着基因扩增,扩增后进行许

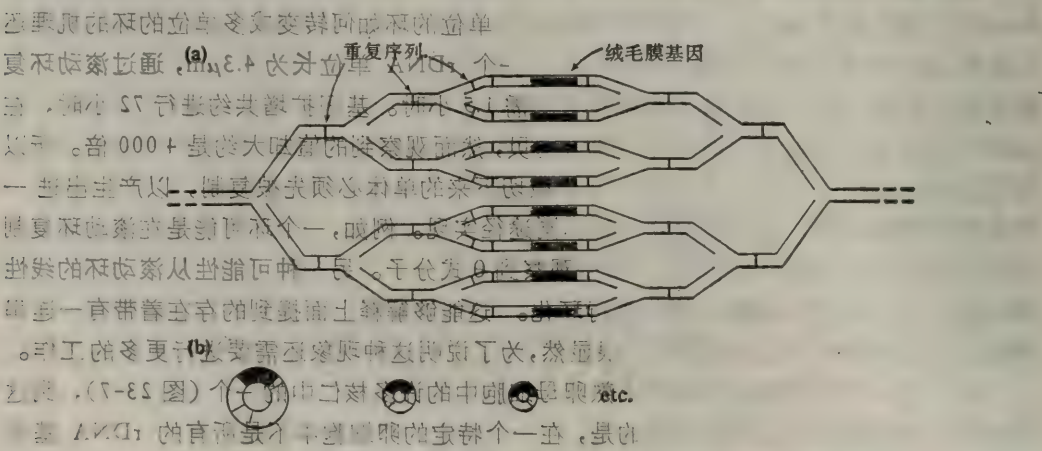


图 23-8 基因扩增作用。(a) 以单个有卵壳蛋白(即绒毛膜蛋白)基因区域(图中有阴影的部分)的一个个的复制起始点起进行三次 DNA 复制的结果。(b) 可能通过在重复序列之间发生重组作用而形成的三种环式分子。



多次细胞分裂,形成特化细胞的克隆。这种体系也利用 DNA 反复地复制来增加基因的拷贝数,不过这些拷贝会被切除,从而形成环状和线性片段。切除过程是由重组序列之间的同源重组所造成。这些环状及片段没有中心粒(见第七章),因此他们在有丝分裂时不能以有序的形式分开。但是,环状及线状片段的数目很大以至随机分离到子细胞中已足够使每个子细胞得到多个拷贝的扩增基因。

### 基因重排: 免疫系统中编码序列之间的连接

一个基因可以通过从远离其启动子的地方移到距其启动子很近的位点而被启动转录,这种基因移动就叫基因重排。第十五章中讨论过的沙门氏菌的相转变 (phase variation) 就是基因位移 (gene relocation) 的一个例子。人们还知道有关真核细胞的一些例子(如酵母中的交配型互变)。但是这些例子过于复杂,本书暂不涉及,而且目前还不知道酵母中的这种位移是否受到调控,也不知道是否受内外原因的影响而改变。在这一部分我们通过有计划地使结构基因的不同片段并联或串联来研究激活基因表达的机理,也就是以免疫球蛋白,即抗体(immunoglobulins, antibodies)的合成为例,研究这种调节机理。

人类能够合成  $10^6$  种以上不同的抗体分子,其“不同”在于一种抗体可对一种不同的靶分子即抗原 (antigen) 起作用。如果各抗体分子被不同的基因编码,那么细胞中必须有很大一部分 DNA 用于抗体合成。然而,如果抗体分子由三种亚基 A、B、C 组成,共有 100 个不同的 A 基因,100 个不同的 B 基因,100 个不同的 C 基因,则所需 DNA 的量将大大减少。这样,  $100 \times 100 \times 100 = 10^6$  种不同蛋白质就可以由 300 个基因编码而成。如果所有这些基因在每个细胞中都有活性,也就是说每个产生抗体的细胞都可合成 300 种不同的蛋白质,那么每个细胞都可以合成出  $10^6$  种不同的抗体。然而已知每一个成熟的产生抗体的细胞都只能合成一种抗体,因此肯定存在着某种对程序化的每个细胞作出反应的机理。

通过这一部分的讨论将可以看到,的确存在着这样的程序化机理。概括地说,胚原基型细胞 (germ-line cell) 具有一整套基因,这套基因的编码序列组合起来就能生成所有可能的抗体分子。在细胞分裂和免疫系统形成的过程中,这套基因的大部分都从子细胞中被切除抛弃了。这种切除是有程序的。在这种情况下成年个体的每一个细胞只能合成一种类型的抗体分子。 $10^6$  种抗体分子中的一种也仅由一个细胞合成,这样,大约共需  $10^6$  个细胞,这个数字只是成年人总数为  $10^{12}$  个的体细胞中的一小部分。抗体与抗原是相互中和的,那么怎样才能产生大量的特定的抗体呢? 这是通过免疫系统对特定抗原所表现的一系列复杂的反应才实现的。在这一反应中,特定抗原能使产生与它对应的抗体的细胞大量地增殖。这种增殖就造成了一个细胞克隆,在这个细胞克隆中的所有细胞都能合成这种特定抗体。

由于免疫系统由许多类型的细胞组成并能对大量的信号发生反应,因而是一个十分复杂的系统。在这一节我们主要关心编码抗体的基因的碱基序列是怎样形成的。这将通过五个部分来阐明 (1) 抗体分子的氨基酸序列的特性是怎样的; (2) 有哪些基因的编码序列,通过怎样的连接才能形成抗体基因; (3) 拼接后的基因的碱基顺序是怎样的; (4) 促使基因拼接的可能信号是什么; (5) 最后的拼接进程如何。然后我们再简单地谈谈怎样指导一个可产生抗体的细胞克隆去合成特定的专一性抗体。

抗体分子氨基酸序列的特点

免疫球蛋白 G(IgG) 是几类免疫球蛋白中的一类,这是一个四聚体分子,由两条轻链(L)和两条重链(H)组成(见第五章)。通过确定许多不同(不同之处在于对不同的抗原发生的反应不同) IgG 分子的氨基酸序列发现轻链和重链都由两个不同的部分组成。其中一部分称为恒定区(constant region)或C区,对于特定一类免疫球蛋白来说,这一区的氨基酸序列几乎不变(只有一个氨基酸可能有变化)。另一区称为可变区(variable region)或V区,任何两个识别不同抗原的 IgG 分子的这一区域的氨基酸序列都不同。L 链和H链上的V区和C区的位置和长度见图 23-9,及本书上册第五章中的图 5-17,5-18,和 5-19。对多数类型的免疫球蛋白来说,C区是用来把抗体分子从产生抗体的细胞中通过其细胞膜运输到血液和淋巴系统,并用于最后对中和的抗原进行处理的过程。V区则负责识别和结合专一性抗原。请注意,在图 23-9 H链和L链的V区尽管氨基酸序列很不同,却有着同样数目的氨基酸残基。

实际上有两种全然不同类型的L链,它们具有相当不同的氨基酸酸序列,被称之为  $\kappa$  链与  $\lambda$  链。它们在功能上有许多相似性,而它们的差别则对于现在的讨论并不重要。我们只叙述  $\kappa$  链的氨基酸序列是如何测定的,对  $\lambda$  链来说,基本的机理是一样的,只在细节上有些区别。

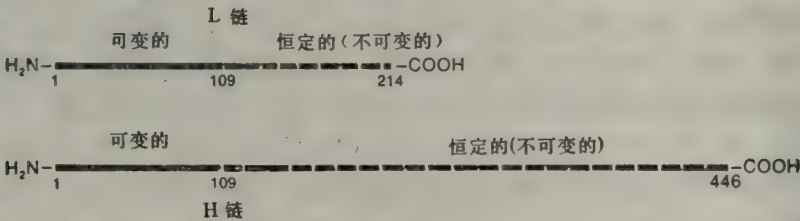


图 23-9 IgG 的L链与H链的V区与C区的长度和位置。图中的数字是指从氨基末端起氨基酸所在的位置。

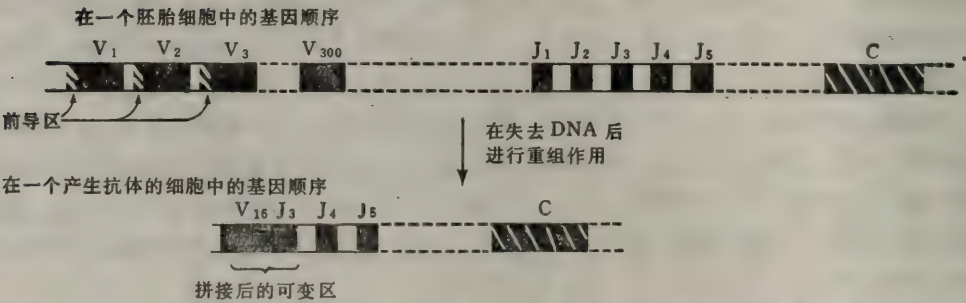


图 23-10 在形成L链可变区时 DNA 的剪接过程。V<sub>16</sub>-J<sub>3</sub> 之间的部分已被移去,推测这是通过位点专一性重组作用实现的。再通过同源重组或前导序列内的位点专一性重组,V<sub>16</sub> 左边的V基因也被除去。

编码序列连接的基因

许多基因可以产生  $\kappa$  型的L链,这些类型有三种类型,即 V、J 和 C。大约有 300 种不同的 V 基因(负责合成可变区 V 区的前 95 个氨基酸),有四种不同的 J 基因(它编码 V



区的最后 12 个氨基酸以及 V 区与 C 区的连接),还有 1 个 C 基因(它编码不变区 C 区)。在胚胎细胞中,V 基因形成一个紧凑的基因簇,在距 V 基因簇很远的地方, J 基因形成第二个紧凑的基因簇,而 C 基因则距在 J 基因簇后面不远处。像图 23-10 所示那样,所有这些 V 基因前面都有一段转录的先导区,而 J 和 C 基因的前面则没有。

### 基因拼接后的序列

编码各种特定 IgG 分子的基因区域已经从各个能产生特定 IgG 分子\*的、众多的小鼠细胞株中克隆(用重组 DNA 技术)。V-J-C 区域也已经从小鼠的胚胎细胞中克隆,这些胚胎细胞还不能合成抗体,因此他们可能含有未发生变化的用于合成抗体的原套基因。用从产生抗体的细胞株中得到的抗体基因的克隆进行对比研究,发现胚胎细胞中出现的很多 DNA 序列在产生抗体的细胞中不见了;而且,缺失的部分总是发生在特定的 V 基因区(这个特定区编码 V 区的前 95 个氨基酸)和 J 基因区(它编码 V 区的后 12 个氨基酸)。这可以通过基因重排来解释。在基因重排中,特定的 V 和 J 基因区域之间的 DNA 序列缺失了,图 23-10 绘出了这样重排的例子。许多编码各种特定 IgG 蛋白的基因序列已被克隆,同时发现在每个克隆中所缺失的是不同的 DNA 片段,例如,在图中  $V_{16}$  与  $J_3$  之间的片段丢失了,而在另一个克隆中缺失也许在  $V_{210}$  与  $J_1$  之间。在这两种情况下,都是一个 V 基因与一个 J 基因拼接在一起形成一个完整的编码可变区的基因。请注意,这是 DNA 剪接,而不是第十一、十五及二十一章所提到的 RNA 拼接。

对于已克隆的 IgG DNA 的碱基序列的研究表明,特定 V 基因与特定 J 基因之间的接合并不总是一样。也就是说,两个并列的 V 基因和 J 基因的终端三联体可以在产生三联体的四个位点中的任何一个位点处发生交换,如图 23-11 对上述陈述作出的解释,其中的接合可能再编码出三种氨基酸,这样就增加了可变区的多变性。

因为 有 300 种不同的 V 基因,4 种不同的 J 基因,连接可以编码 2.5 种(大致平均)不同的氨基酸,这样就可能有  $300 \times 4 \times 2.5 = 3000$  种不同的可变区。H 链基因的结构与 L 的很类似,但不完全相同(有四种不同类型的基因),所以,这样在 H 链中就约有 5000 种不同的可变区。每个 IgG 分子都具有两个相同的 H 链和两条相同的 L 链,所以,由  $V_L$ 、 $V_J$ 、 $V_H$  和  $J_H$  基因大约可形成  $3000 \times 5000 = 1.5 \times 10^7$  种不同的 IgG 分子。这可以充分地说明抗体分子的多样性。

$5 \times 10^7$  种不同的抗体专一性这个数目实际上是个保守的估计,因为还有其他方式能造成抗体多样性。在生成抗体的细胞的发育过程中,V 区对于很高的体细胞突变率是敏感的。因为 有这些突变,即使不同的细胞克隆产生的抗体有完全相同的 V-J 连接方式,这些克隆产生的多肽的序列也可以不同。

### 基因拼接的可能信号

上述 V 基因及 C 基因的剪接,人们认为是通过一种位点专一性重组的机理而进行的。

---

\* 这些细胞株是从称为骨髓瘤 (myelomas) 的致癌肿瘤衍生而来。骨髓瘤是一种肿瘤,在这种肿瘤中单个细胞可以程序化地形成一种抗体,将这种单个细胞进行繁殖,能产生出一个很大的细胞克隆。这种细胞克隆中的每一个细胞都只产生该种抗体。人们可以分离到单个的这种细胞,单个的这种细胞在细胞培养液中可以无限地生长;细胞分泌大量的纯抗体,从细胞培养液中可以回收到此种纯抗体。

引起 V-J 连接的信号很可能分别在 V 基因右边和 J 基因左边的一对专一性的碱基序列中。

先看与 V 基因相关的序列:

V 基因-CACAGTG-11 碱基-ACAT AAAC

以上是一个 7 碱基对的回文结构序列、一个 11 碱基对的间隔区, 和一个富含 (A + T) 的 8 碱基对的序列, 组合在一起紧接在每个 V 基因的右侧。

下面再看看每个 J 基因左侧的序列:

GTTTTTGT-22 碱基-CACTGTG-J 基因

这个序列与紧靠着 V 基因的序列相同(除去第四个碱基), 也是一个碱基对的回文结构序列(从右向左读), 然后接着一个 22 碱基对的间隔区和一个富含 (A + T) 8 碱基对的序列。这个 8 碱基对的序列几乎与 V 基因附近的那个 8 碱基对序列互补。推测, 这些序列可能被剪接体系识别。目前, 无论是对于拼接所需要的酶, 或是对于化学交换过程的性质以及触动拼接的化学事件都不了解。

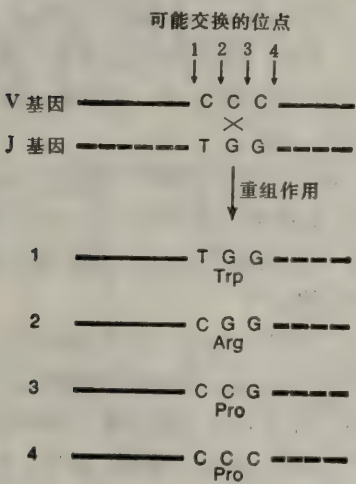


图23-11 在V-J 连接处四种可能的连接方式, 由此可能增加三种不同的氨基酸。

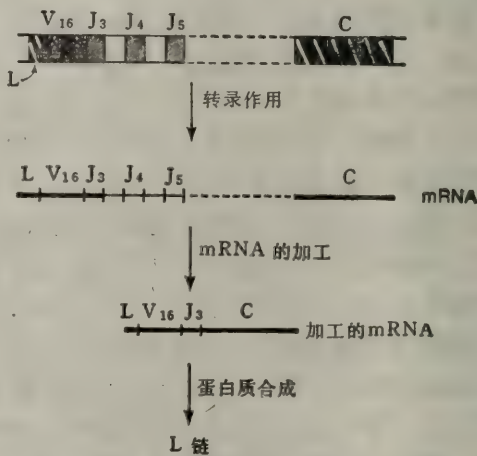


图23-12 一个特定 IgG 分子的 L 链的产生。mRNA 中粗线区域代表用来产生最终 L 链的编码序列。

最后的拼接事件: RNA 的拼接

图 23-12 指出 DNA 拼接并不能产生最终的 L 链, 这是因为: (1) J 基因与 C 基因之间还保留着间隔序列; (2) 实际上最终编码 L 链的只是一个 V 基因和一个 J 基因。而拼接后的 DNA 通常仍含有很多 V 基因和 J 基因。正确的氨基酸序列是通过最后的 RNA 拼接才获得的, 如图中所示。应当注意特定的 V-J 接头是怎样决定 RNA 拼接形式的, 例如, 总是除去前导序列与 V 基因之间的 RNA 片段, 又总是除去 C 基因与特定 V-J 接头中 J 基因右端之间的 RNA 片段。

H 链的多样性

产生 H 链的氨基酸序列多样性的机理与 L 链的类似, 虽然知道的还不详细。H 链也



具有许多 V 与 J 基因，但是 H 链在两方面比 L 链复杂：(1) 在 V 和 J 基因之间增加了许多大套基因。称作 D 基因，所以可能有很多 V-D，D-J 接头；(2) 有八个不同的不变区 (C 区)，C 区的选择决定了免疫球蛋白的特定种类 (如 IgD，IgG，IgM 等等)。很有意思的是上面所说的启动 L 链中 V-J 连接的 7 碱基对的迴文结构也出现在 C 链体系中的相应位置上。

产生抗体的细胞的双倍体

由于产生抗体的细胞是双倍体，因此出现了一个很有意思的问题：一种特定的细胞仅仅产生一种 IgG 分子，那么，或者是两条染色体上发生同样的重排事件，或者是两个 V-J-C 区中的一个保持沉默，到底是哪一种？现已得到的证据表明，在从胚胎细胞到形成能够产生抗体的细胞的发育过程期间一个完整的有功能的基因一旦装配完毕，基因重排就立即停止。而那些仍然处于胚胎构型的基因或未完成拼接的可变区中的基因，则或者是不能表达，或者它们的表达产物是没有功能的。

程序性

一个重要的问题是，DNA 重组是否就发生在对某一特定抗原作出反应时？最好的证据指出这些重组事件随机地在发育过程中发生着。当分化过程完成之后，大约形成了  $1.5 \times 10^{-7}$  种不同类型的产生抗体的细胞，其中的每一种细胞都能合成出一种抗体。那么一个机体该在什么时候大量地产生出某种抗体以应答特定的抗原呢？对于这个问题的答案还不肯定。但下面被称作克隆选择 (clonal selection) 的机理至少从大体上说很可能是正确的。每个产生抗体的细胞产生少量的专一性抗体。这种抗体中的某些结合到细胞表面。当这样的细胞遇到一种能与这种专一性抗体反应的抗原时，就形成一个复杂的抗原-抗体复合网，因为每个抗体可以和两个抗原分子相互作用，而每一个抗原又能与两个抗体分子结合 (见第六章)。这些反应与细胞膜表面的抗体分子的耗能运动偶联，促使形成一个称为细胞帽 (cell cap) 的抗原-抗体网 (图 23-13)，然后细胞膜向内折叠包围

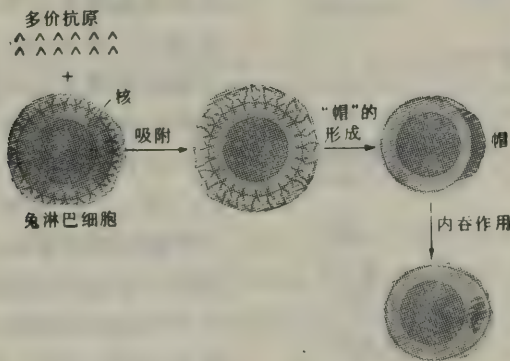


图 23-13 帽化现象。兔淋巴细胞产生均匀分布于细胞表面上的抗体 IgG，将这种细胞暴露于抗原中，于是，抗原与抗体分子互相结合。这个过程是一个需能过程，但对此还不甚了解，只知道是 IgG 分子向抗原分子靠拢。如果抗原是多价的 (有多个结合位点)，就形成一种被称为帽子的复合物。通过细胞的内吞作用，帽子被吸收到细胞内。在这个过程之后，细胞程序化地产生大量抗体。若将抗原与荧光素偶联，那么用荧光显微镜可以看到荧光帽；若抗原是有放射性的，那么在放射自显影照片上每一个细胞的一个区域都有一块是黑色的。

这个帽子，进而这个帽子被吞入细胞(这个过程称为细胞的内吞作用，endocytosis)。然后，在一个总的说来是模糊不清的方式下，促进着细胞分裂，接着广泛地产生抗体。于是，就产生了一个能合成大量的专一性抗体的细胞克隆。

## 初级转录产物合成的调控

用调节转录的方法可以控制很多基因产物的合成。这多半是低拷贝数和中等拷贝数的 mRNA 编码蛋白质的主要机理；这一机理也同样适用于多拷贝的 mRNA。在这一节中我们将讨论几个用不同方式开启和关闭转录的例子。首先讨论真核基因排列即真核基因构造 (organization) 的一些特征，从而阐明真核细胞的调节不同于原核细胞操纵子的那种直接调节机理。

### 真核基因构造影响调控的特征

原核细胞操纵子模式以及它的许多变种的一大特征是：某些调节分子作用到与每个启动子相邻的一段 DNA 序列上，这种作用决定转录是否发生。调节位置总是与启动子相邻。原核生物的这种调节机理的简单性使得这种调节位置不可能与启动子离开很远。事实上，由于20多年来我们对于细菌操纵子的研究与认识，使得我们很难转而理解真核细胞的基因表达的调控，因为在真核系统中，调节位置极少邻近于启动子。

真核基因调控中的另一个重要问题，是解释在 DNA 上相距甚远(甚至在不同的染色体上)的一组基因是怎样地能够同时被启动或被关闭的。有人提出一组基因中的每一个

基因都被同样的调节序列引导着，所以一个简单的调节因子能使这套基因中的所有基因同时活化或同时失活。这种机理与前面(第十四章)提到的原核基因的调节相类似。然而，真核体系还受一些本身亦由某些信号开动的中间调节因子的调节。图 23-14 示出可能发生的调节过程的概要途径。真核调节的另一个普遍存在的性质，是有一些基因产物为了要对几种不同的信号作出反应而合成的。例如，两种不同的激素可以诱导同一种蛋白质的合成。更有甚者，自然界中有一些组合式基因可以被不同的信号以不同的结合所激活。例如，一个信号能激活 P 和 Q 基因，另一个信号激活 Q 和 R 基因，第三个信号激活 P 和 R 基因。对这些现象可以用以下假设加以说明。第一种：有多组分的信号分子(复合的激活基因)及一些单个的受体基因，复合激活基因中的一个激活基因产物结合到一个受体

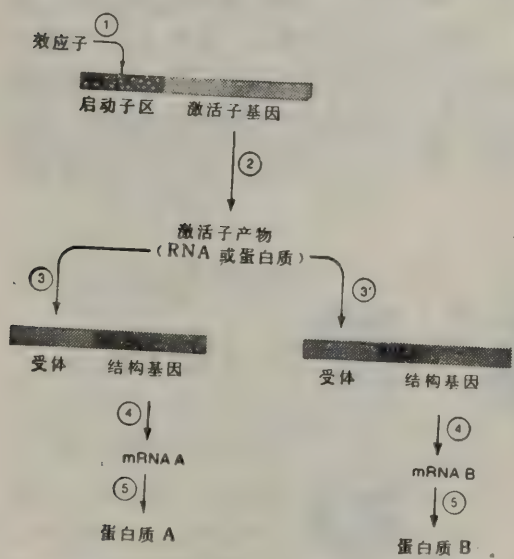


图 23-14 真核基因调控的一种可能的简单方式。每一步都按其发生顺序标号，受体基因在第3步或是被 RNA 或是被蛋白质激活。这个过于简单的模型很可能并不普遍适用。

活基因)及一些单个的受体基因，复合激活基因中的一个激活基因产物结合到一个受体



基因上,打开一个结构基因;第二种: 有两个(或多个)受体的基因(复合的受体基因)及一些结构基因,激活基因可以结合到复合受体基因上;只要每一个复合的受体基因中的一个受体基因受到激活,紧接的结构基因就能被活化。关于这些体系是如何工作的,在图23-14中进行了介绍。

正如我们已经申明的,图 23-13 及图 23-14 模型都肯定地过于简单化(未达到用分子术语解释调控)。但对于读者来说,重要之处就是只要明了相隔甚远的基因是怎样被协同调节的,而且要知道调节的几个阶段的要求是什么。下面的模型主要为思考真核调控作用的某些性质提供了一个简要的框架。

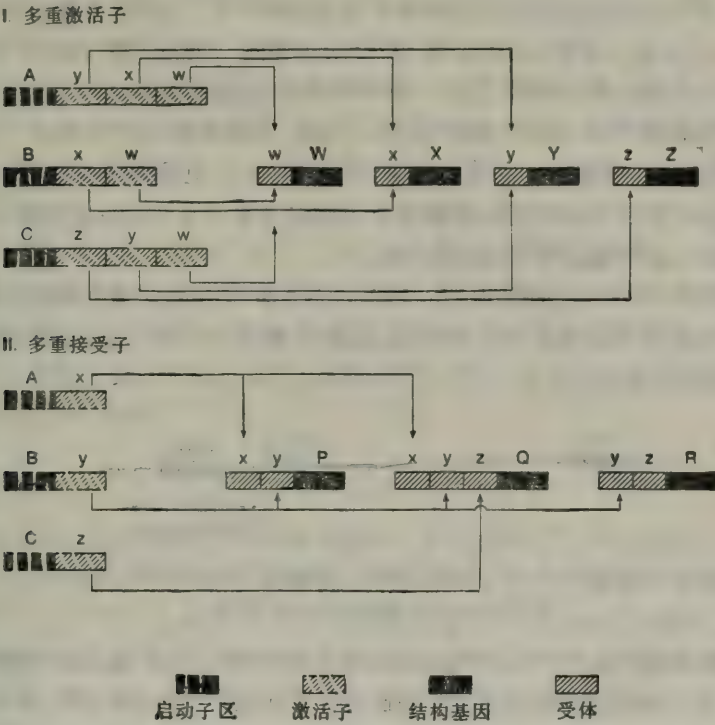


图 23-15 受多重调节的基因家族的可能结构。上面一行 (I) 中,具有复合激活基因的结构单位可以激活几个结构基因。下面一行 (II) 中,一个单独的激活基因(或称激活子)可以打开几个结构基因,这几个结构基因有着共同的复合受体基因(或称接受子)。

### 酵母中的调节单位

作为非寄生的单细胞有机体的酵母,必须能够像细菌那样对环境的改变被动地作出快速的应变反应。酵母中以此种方式起作用的一组基因是负责半乳糖代谢的 GAL 体系。这一体系由位于几个染色体上的许多个基因组成。在这里我们将只考虑 GAL1 基因的调节,图 23-16 显示了它的调节区域。它由一个上游激活位点 (UAS<sub>G</sub>)、一个 250 个碱基对的间隔区、TATA 盒以及 GAL1 基因按以上顺序排列组成。在另一个染色体上只有一个称为 GAL4 的基因,它的基因产物是一个正调控物质,这种物质可因结合半乳糖而被激活。激活了的 GAL4 蛋白质结合到 UAS<sub>G</sub> 中的一段 15 个碱基对的序列上,开启位于 TATA 盒下游几个碱基对的一个位点,于是开始进行转录。UAS<sub>G</sub> 和 GAL1、TATA

盒间的距离并不很严格,例如,用遗传工程的方法将其向上游移去 500 个碱基对,而不影响 GAL1 基因的可诱导性。显然,  $UAS_G$  多半是转录的一个一般激活物。其证据是如果将它放到另一个酵母基因 CYC1 的 TATA 盒上游的几百个碱基对处, CYC1 基因也可以因此被半乳糖诱导出转录作用。

对位于  $UAS_G$  和 GAL1、TATA 盒之间的间隔区进行基因操作已使我们或多或少地了解了这两段序列之间的关系。人们推想,由于被半乳糖激活了的 GAL4 蛋白质与  $UAS_G$  结合促进了转录作用,所以 RNA 聚合酶应当结合到与激活了的  $UAS_G$  相距尽可能近的地方,以便与 GAL4 蛋白接触。这种推想是合理的,因为至少这是用原核生物调控方式去思考而得出的看法。然而, TATA 盒是形成“开”型启动子复合物 (open-promotor complex) 的位点, RNA 聚合酶 II 不可能跨越  $UAS_G$  和 TATA 之间的距离,这与原核系统中 RNA 聚合酶可以同时与 -35 序列区以及 Pribnow 盒区相接触相反。有三种简单的方式使得 RNA 聚合酶可以从  $UAS_G$  到达(接触) TATA:

1. DNA 可弯曲,使得  $UAS_G$  和 TATA 互相毗邻。但是由于弯曲很可能发生在一段独特的序列上,以及  $UAS_G$  可能被重新放置到或者是 TATA 处或者是另一个结构基因 (CYC1) 附近,这一想法看来又不太可能。

2. 将被激活的 GAL4 蛋白结合到  $UAS_G$  处,这样可以改变染色质的结构。这一结构上的改变又可能向下游波及到 TATA 区,人们已假设 TATA 区在这个模型中是 RNA 聚合酶的唯一结合位点。



图 23-16 酵母的 GAL1 区域。GAL4 基因产物与结合着的半乳糖,作用于  $UAS_G$ , 结果合成出 GAL1 RNA。

3. RNA 聚合酶可沿 DNA 从  $UAS_G$  滑动到 TATA 区。人们通过在  $UAS_G$  和 TATA 之间插入一个转录终止位点而对第三种假设进行了研究。这一插入取消了半乳糖对 GAL1 转录的诱导。由于 RNA 聚合酶可能在终止位点脱落,所以该结果支持滑动假说。然而,由于一个终止蛋白可以与终止位点结合妨碍理论上假设的染色质结构发生波及 TATA 的改变。这样看来,第二种解释也有可能;但是争论的焦点是,在真核细胞中还没有检测到有终止蛋白的存在。

第二个实验同样显示了干扰  $UAS_G$  和 TATA 之间的间隔区可能会干扰基因表达。在这个实验中,曾把大肠杆菌 *lexA* 的操纵基因插入到  $UAS_G$  和 TATA 之间,此举对 GAL1 的可诱导性并无影响。然而,如果把 *lexA* 基因也插入其中,可以合成出 LexA 蛋白质,那么 GAL1 的转录则就受到抑制。如果 *lexA* 的操纵基因位于  $UAS_G$  的上游,这个占据位置的 *lexA* 操纵基因则并无抑制效应。这些结果表明  $UAS_G$  和 TATA 之间的结合蛋白阻挠转录的启始,这与上述的假设 2 和假设 3 是一致的。

对另一个酵母体系(配对型互变 mating-type interconversion)的观察提示,有时染色质的改变可起关键性的作用。在这个体系中,从  $UAS$  上游起的几千个碱基对的区域是称为沉默子 S (silencer S) 的位点。那些被称 SIR 的几个基因的产物与此位点相



互作用。当 S 被它们占据时, S 范围内几千个碱基对中的几个启动子便失去活性。这个结果可解释为 SIR 诱导染色质结构发生变化阻断了转录作用。

合成异细胞色素 c 的 *CYC1* 基因有两个上游位点 *UAS<sub>c1</sub>* 和 *UAS<sub>c2</sub>*, 分别位于距转录起始点 -275 和 -225 的地方。这两个位点的活性都由血红素(细胞色素的辅助因子)调控。基因 *HAP1* 和 *HAP2* 编码专一性的正调节蛋白因子, 这两种蛋白质都与血红素结合, 而且一旦血红素结合到以上两个位点后, 这两种物质就分别地与各自的 *UAS* 互相作用。每个 *UAS* 在功能上是分别发挥作用的, 而且它们的效应是相加性的。具有串联(一前一后)性 *UAS* 的酵母可以对特殊的调节性刺激作出反应。在这个例子中, 不同的信号分子决定了合成异细胞色素 c 的量, 即量 1, 量 2, 或者是量 1 + 2。

上述例子是关于正调节体系的。酵母也有一个负调节体系。两个几乎相同的基因 *ADRI* 和 *ADRII* 编码着乙醇脱氢酶。*ADRI* 基因的转录是组成型(constitutive)的, 但 *ADRII* 基因的转录受葡萄糖抑制。位于 *ADRII* 上游的一个调控区与葡萄糖的抑制有关: 若这个区域缺失, 便发生对葡萄糖不敏感的组成型转录。这个位点与至今已讨论过的 *UAS* 的不同之处, 在于它与 *TATA* 之间的距离是很严格的。例如, 若有一个插入序列置换掉它的上游区, 则失去葡萄糖的抑制效应。还有, 这个区域不是基因专一性的, 因为如果用遗传工程的方法移去 *ADRI* 的上游, 那么 *ADRI* 基因的转录也不再是组成型的, 而成为可以被葡萄糖抑制的。这个区域的行为类似于原核的操纵基因, 但是至今还没有探测到相对应的阻遏物。

### 激素对转录的调控

在已知的真核细胞的基因活性的调节物中, 研究得最广泛深入的是一些激素。这些细胞外物质, 或者是小分子, 多肽, 或者是小的蛋白质, 由产生激素的细胞运送到靶细胞上。进一步人们还发现一些激素, 如许多类固醇激素(例如雌激素、孕激素、醛固酮、糖皮质激素和雄激素), 以及一般的代谢性激素(例如胰岛素)。它们频繁的调控作用是通过开启转录而实现的。若一种激素可以调节转录, 则它必须多少能向 DNA 发出信号。激素穿入到一个靶细胞中并运送到核的过程很复杂, 目前也只了解其概况, 如图 23-17 所示。固醇类激素(H)是疏水性分子, 可自由地出入细胞膜。一个靶细胞具有一个专一性的细

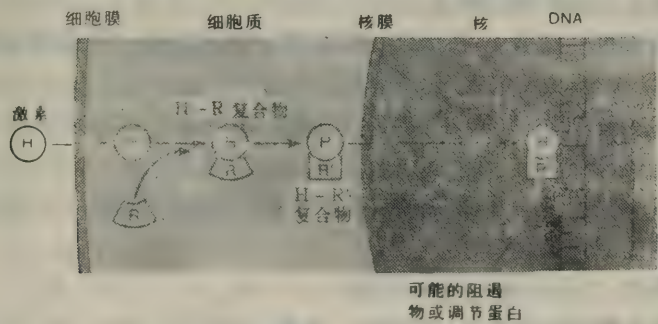


图 23-17 表示激素 H 是如何通过结合到细胞质的受体而达到 DNA 并触发转录的图解。一种调节蛋白可能可以防止 H-R' 复合物结合到启动子(P)上, 或可以促进它结合到启动子上。

胞质受体,它可与激素形成复合物(H-R)。受体R在H-R复合物形成后通常受到一定的修饰(构象的或是化学组成的);修饰过的受体以R'表示。然后H-R'复合物通过核膜进入核。在核内,无论是H-R'复合物或是单独的激素均与染色质结合,并使DNA开始转录。虽然在某些体系中复合物可能直接地结合到DNA上,但是一般来说结合很可能不是直接的,而是通过专一性的DNA结合蛋白而实现的。

一个研究得比较透彻的激素对转录的诱导作用的例子,是雌激素(estrogen)的刺激导致鸡输卵管中卵白蛋白(ovalbumin)的合成。将雌激素注入鸡体内时,输卵管组织便开始合成卵白蛋白的mRNA。只要一直给予雌激素或雌激素样激素(例如二乙基己烯雌酚(diethylstilbesterol)),该合成便继续下去。一旦停止供给激素,合成的速度便下降。无论是给激素前或是停给激素后60小时,都不能检测到卵白蛋白的mRNA。在解释这个实验现象时必须注意避免由测量转录的方法而带来的假像以及内含子的存在。专一性的mRNA,例如卵白蛋白mRNA,是由标记了的mRNA与含cDNA的质粒经分子杂交而检测到的。假设一个初级转录产物有大量的大内含子,这也许会干扰特定mRNA与cDNA的杂交。如果与加工过的mRNA进行比较,初级转录子与cDNA结合得很少,那么用这种方法检测mRNA的能力也许只能说明加工过程,而不能表明有没有RNA的从头起始(de novo)的合成过程。另有两种直接的方法已指出激素的作用是促进转录而不是加工mRNA。一种方法是显示新合成的RNA中有内含子,这可以通过在杂交试验中使用天然基因的含有内含子的cDNA的克隆而达到。另一种方法是测量出现的RNA的大小。如果转录是由激素开启的,那么应当可以检测到比加工过的mRNA更大些的初级转录产物,而且应当在比加工过的mRNA出现更早些的时候观察到这种初级转录产物。但是如果仅仅是促进加工过程,那么被检测到的大RNA应该没有量的变化,而只是较小的、加工过的mRNA在浓度上有所增加而已。现在,正是这个实验指出了雌激素的作用是前者,它的确促进了卵白蛋白初级转录产物的转录过程。

当给鸡注射雌激素时,只有输卵管合成mRNA;其他组织因为缺乏细胞质的激素受体,对激素的刺激并不敏感。这种类型的缺陷(指缺乏受体)毫无疑问是某些组织对某种特殊激素不敏感的原因。至于受体是怎样在某些细胞中合成,而不在所有细胞中合成的原因则所知甚少。这是另一个问题,即分化问题。

到目前为止我们还未讨论激素是怎样开启转录的,但是我们至少可以检查五种可能的基本机理而避免去讨论复杂的细节。

1. 激素直接与DNA结合,并促进RNA聚合酶或促进一种在聚合酶在与DNA相互作用时必不可少的蛋白质(转录因子)结合到DNA上。

2. 激素与一效应蛋白质(例如lac操纵子中的CRP蛋白质)的结合。这种效应蛋白质结合到DNA上处于RNA聚合酶结合位点上游的位点上。

3. 已经与DNA结合的一个蛋白质被激素激活,并且这种激活了的形式是RNA聚合酶结合时所需要的。

4. 激素是诱导物,它可以使阻遏物失活。

5. 激素引起染色质的结构改变,把DNA暴露出来以使RNA聚合酶能结合上去。

在前三种机理中,转录受到正调节,在第4和第5种机理中,转录受到负调节。目前人们只研究了几种受激素诱导的体系,并且对它们还没有清楚的认识。



## 由转录因子介导的调控

在第十一章我们已经看到,在很多真核体系中转录的起始需要有 DNA、RNA 聚合酶以及一个或多个称作为转录因子的蛋白质所形成的转录复合物。随着更多体系得到研究,很清楚地表明,调控经常受到转录因子的作用,并且也已观察到两种机理:(1) 转录复合物形成过程受到调控作用;(2) 通过一个小分子的促进,RNA 聚合酶结合到 DNA 上的过程受到早已存在的转录复合物的调控。在这一节里,我们将探讨具有这种效应的几个例子。

非洲爪蟾基因组具有两组编码 5S rRNA 的多基因家族(见图 23-1),它们被称作卵母细胞型家族和体细胞型家族。占 2% 的 5S rRNA 基因的体细胞型基因家族担负着体细胞中 95% 以上的 5S rRNA 的合成,而在卵母细胞中的卵母细胞型基因家族只有初级转录过程。以上这种不同的基因表达方式是因为在内在的调控区中转录复合物的形成受到不同的调控所致。若把染色质从体细胞中分离出来,可以发现体细胞型 5S rRNA 基因是活泼的转录复合物中的一部分。这一复合物包含一个称为 40-kd 转录因子的蛋白质。然而卵母细胞型基因则没有这种与它结合的转录因子。而且事实上,含有组蛋白 H1 的某种结构阻止它与转录因子的结合。这种差异亦可在体外重复。如果体细胞的染色质在浓盐溶液中被洗去一个 40-kd 蛋白质,那么 5S rRNA 基因的转录将消失。再加入 40-kd 蛋白因子后又恢复转录能力。若在加入 40-kd 因子之前就把 H1 加入,则将阻碍转录复合物的形成,但若先加 40-kd 因子则不阻碍。注意 H1 具有原核细胞中阻遏物的特征。

40kd 蛋白因子在胚胎发育中的作用尚未得到完全的阐明,但人们认为以下种种事件的顺序是至关重要的。每个非洲爪蟾细胞中包含有大约 800 个体细胞型 5S rRNA 和 40 000 个卵母细胞型 5S rRNA 基因。它们两者的内在调控区都具有 50 个碱基对,但是有两个碱基对是不同的。人们认为这个差别大大地降低了 40-kd 蛋白质对卵母细胞基因的亲和性。幼年卵母细胞包含有过量的 40-kd 蛋白质,于是所有的内在调控区都被激活。但是,由于某些目前还不清楚的原因,40-kd 蛋白因子的量在晚期的卵子发生中以及在减数分裂中大大地下降了。受精之后以及在胚胎发育期间 40-kd 蛋白因子总量是恒定的,但是,在细胞分裂后,它在每一个细胞中的含量下降了。在这个时期内两种 5S rRNA 都不合成。尽管原因还不清楚,但人们提出了种种假设。例如缺乏足够量的 RNA 聚合酶 III,在转录复合物中,以及在核小体的复合物中都缺乏所需要的其他的蛋白因子。在具有 4 000 个细胞的胚胎阶段胚胎细胞以任意的速度开始合成着 5S rRNA。在这时,每个 5S rRNA 基因有大约 10 个 40-kd 蛋白因子分子,即因子与基因的比例是 10。而且卵母细胞型和体细胞型的 5S rRNA 基因都被转录。经过 3—4 次细胞分裂后,因子与基因的比例下降,接近于 1,卵母细胞型 5S rRNA 基因的转录完全被阻遏了。推测,在染色质复制期间 DNA-蛋白质复合物发生解离。而且,由于 40-kd 因子的浓度低,H1 的阻遏性结合就占了优势。从这以后,40kd 因子就随细胞分裂而合成,保持大约每个 5S rRNA 基因有 0.25 个 40-kd 因子分子这个比例。这样,在体细胞中就因为 40kd 因子分子比 5S rRNA 基因少,卵母细胞型 5S rRNA 基因便不转录。推测,40-kd 的转录蛋白因子对于体细胞基因的内在控制区的高度亲合性可能使体细胞型基因得以优先表达。

组蛋白基因的表达也是由转录因子介导的。只有当 DNA 复制时才需要新合成的组蛋白,以形成染色质(第八章)。对于小白鼠,在细胞周期的 DNA 复制相中各种组蛋白的转录的确增加了 20 倍。在组蛋白 H4 基因的转录起始点上游 70—110 碱基对处有一个启动子。只有当一个特定的转录因子结合到这个启动子上后,H4 基因才发生转录。因此,H4 基因的转录作用是受到一定的调节的;这种调节或者是通过激活转录因子,或者是通过供给转录因子而实现。在这里,后一种方式的可能性更大。这是因为有初步材料证明,在转录周期的前期,转录因子的浓度有所增加。因此,H4 基因的调控不是直接的,而是通过激活转录因子的合成,提供可使用的转录因子而实现的(图 23-13)。目前还不知道这种激活作用过程的机理。因为在同一时期中,所有组蛋白(H2A, H2B 等)或是全需要,或是全不需要,所以人们很容易想到所有的组蛋白基因会使用同样的转录因子。然而进化却选择了另一条途径:小白鼠的 H2A 和 H3 基因都有其自己的转录因子。各个转录因子的合成则很可能由同样的信号控制,但是还没有肯定的证据。

在酵母中,组蛋白基因的表达则有着另一种要求(小鼠没有这种要求)。通过基因工程手段可以分别移走一些序列,使基因与一个自然的 DNA 复制起点之间的相对距离依次不同,然后再测试基因所发生的不同表现。结果发现,只有在这些基因与复制起点相距甚近时基因才能表达。很可能是转录因子结合到 DNA 上要求一定的条件:只在复制起始点使那些与起始点相联系的染色质的结构发生改变时才能结合到启动子上。但是这一点还纯属推理,人们对此还有许多怀疑。

转录因子调控的另一个著名的体系是果蝇的热休克族基因。这正如第十五章中所阐明的,大肠杆菌的热休克调控子那样,很多生物在被加温到最适温度范围以上时,受到热的压力,诱导合成出称之为热休克蛋白(heat shock protein)的一系列蛋白质。无论在细菌中还是在高等真核生物中,热休克基因都在整个基因组中散在地分布着。对果蝇的 Hsp70 蛋白有过较详细的研究,使果蝇受热能引起 Hsp70 mRNA 分子数增加 1000 倍。Hsp70 基因的 TATA 盒位于转录起始点上游 10 碱基对处。再往上游 60 碱基对处有一个称之为 HSE(Heat Shock regulatory Element)的调控序列。热休克转录因子(HSTF)能结合到这个序列上,以激发启动转录。继续往上游 800 碱基对还有另一个 HSE 序列。

要理解“热”对于 Hsp70 mRNA 的诱导,就必须了解热效应的作用。有两种可能性:(1)在通常情况下,HSTF 结合到接近于 HSE 序列的位置上,但是这种结合作用太微弱,不能推动转录。在加温之后,HSTF 发生构象变化,形成可紧密结合的构象,从而使转录开始。(2)HSTF 通常与较远的位置上的 HSE 序列微弱地结合;加温使 HSTF 从这个位点解离下来;或者游离的 HSTF 分子和与之较近的 HSE 序列随机无序地碰撞,或者是 HSTF 沿 DNA 滑动,直到 HSTF 接近于 HSE 时与 HSE 结合,从而使转录起始。但是对于这两种可能性目前还没有实验数据去区分。

人们在  $\beta$  珠蛋白区域还发现了一个有趣的现象。从出生 9 天的小鸡的红细胞中分离出的  $\beta$  珠蛋白的 DNA,在基因表达时发现其上有一个超敏感位点(hypersensitive site)(第十二章)。而在更小的鸡中, $\beta$  珠蛋白基因不表达,也没有发现这个超敏感位点。从 9 天的鸡细胞中分离出了一种蛋白质,它可以诱导出超敏感位点。如果把这种蛋白质加到从更小的鸡细胞中提取到的染色质上,则在这些染色质上也出现超敏感位点。这个蛋白质就结合到受切割区域中的专一性序列上,即超敏感位点上。可以假设:这种蛋白质



与 DNA 的结合取代了组蛋白与 DNA 的结合,或者至少是减弱了组蛋白与 DNA 的相互作用,从而改变了这个位点处的染色质的结构,使 DNA 易于受核酸酶的降解。

以上所述的各个体系是通过控制转录因子的结合来实现调控的例子。初步的证据还提供了由小分子物质来激活已结合在 DNA 上相应位点的转录因子的例子。在这个例子中,只有当一个激素-受体复合物结合到转录因子上后,DNA 聚合酶才能结合到启动子上,启动转录作用。

增强子活性的调节

在启动子上游或下游几百个碱基对处的增强子 (enhancer) 可以显著地增加启动子的转录速度(见第十二章)。尽管增强子的工作机理还不清楚,但是可以想象控制增强子的活性也可以作为一种调控方式。人们通过实验手段,可以把增强子移到整个基因组中任一位置,它就会对附近的任何基因的表达发挥增强效应。因此,增强子对其起作用的基因没有专一性。另一方面,增强子经常表现出种的专一性和细胞类型的专一性。例如,小白鼠的免疫球蛋白的H链增强子在骨髓瘤细胞(这种细胞通常合成免疫球蛋白)中是有活性的,但在成纤维细胞(它通常不合成免疫球蛋白)中则无活性。这说明转录因子可能是一种专一性蛋白质,增强子增强活力时需要它。从人的 HeLa 细胞中已经分离到了这样的一个为 SV40 和多瘤病毒增强子所必需的转录因子。最近(1985 年)的一个实验表明,这个多瘤病毒增强子的活力可能被抑制。又有人制备了两个质粒:(I) 一个质粒含有多瘤病毒增强子和人的  $\beta$  球蛋白基因和 (II) 另一个质粒含有腺病毒的 E1A 调控基因(见第二十三章真核病毒)。当质粒 I 进入 HeLa 细胞后, $\beta$  球蛋白基因活跃地进行转录。然而,当两个质粒同在细胞中时,只能合成出很少的  $\beta$  球蛋白 RNA。另外几个实验证明,E1A 蛋白就是这个增强子的抑制物。但目前还不清楚究竟是 E1A 蛋白直接地结合到增强子

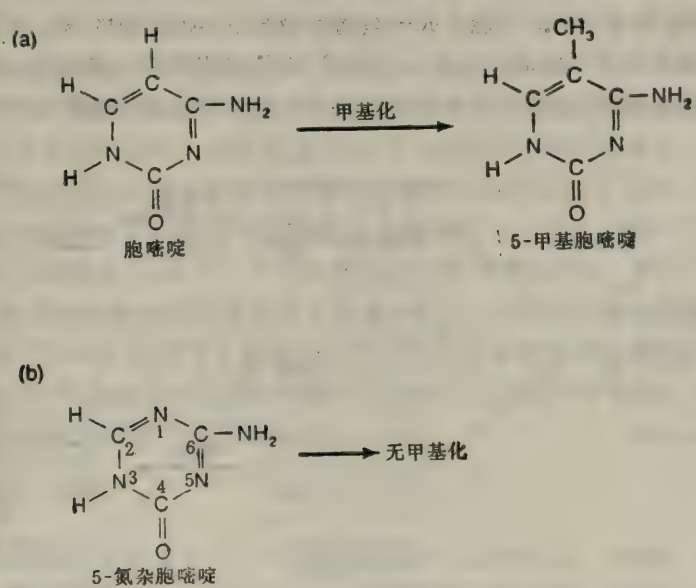


图 23-18 (a) 胞嘧啶经甲基化后形成5-甲基胞嘧啶。(b) 5-氮杂胞嘧啶, 示出一个 N 原子取代了胞嘧啶中的可被甲基化的那个 C 原子。在 5-杂氮胞苷中右下角的 N 原子连着一个核糖。

上，还是通过一些由 EIA 蛋白促进合成的产物来抑制。考虑到在腺病毒感染的细胞中 EIA 蛋白通常对细胞基因和病毒基因的转录都有促进作用，因此，后一种说法更有可能。

正常的细胞是无 EIA 基因又无多瘤病毒增强子的，因此前面介绍的是一个完全的人造系统。尽管如此，以上结果还是说明了增强子能被失活，因而很可能存在着增强子的阻遏物，不过目前尚无直接证据支持这种观点。

甲基化作用

多数细胞都有几种 DNA 甲基化酶。有一些就是限制性酶的甲基化活性，这是用来使限制性内切酶的作用位点甲基化以免自身 DNA 受到破坏。另一类限制性酶在特定的 CpG 序列处使胞嘧啶甲基化(图 23-18)。在少数情况下，尤其是在脊椎动物中，特定的 CpG 序列的甲基化能够阻止某些基因的转录。有人提出这种甲基化可能对真核细胞的调控和发育有普遍意义，然而对这一论点还有许多不同意见。这个论点遇到的主要困难是在无脊椎动物中很少发生甲基化作用，而在昆虫中则根本不发生甲基化，即使在哺乳动物中关于甲基化对转录的作用也有许多未见到效果的例子。以下是一些甲基化作用后产生调控效果的证据：

- 1. 某些基因在不表达这些基因的细胞中高度甲基化，而在表达这些基因的细胞中不发生甲基化。
- 2. 在体外使克隆了的  $\gamma$  球蛋白基因的上游位点(即此基因的 5' 端)甲基化，接着用  $\text{CaCl}_2$  将经过这种处理的 DNA 转化到敏感细胞中，转录便受到阻止。如果甲基化作用发生在基因上游位点以外的位置，则转录不受抑制。
- 3. 如果在培养基中加入 5-氮杂胞嘧啶，新合成的 DNA 中就含有这个碱基类似物。含有这个类似物的 CpG 位点是不能发生甲基化的(5-氮杂胞嘧啶是所有甲基化的普遍抑制剂)。在这种培养基中的细胞能生成一些通常情况下不能合成的蛋白质。
- 4. 所谓的看家基因 (house keeping gene) 就是那些维持一般细胞正常功能所必须的而且始终继续被转录的基因。很少在这些基因的起始区或起始区附近发生甲基化作用。

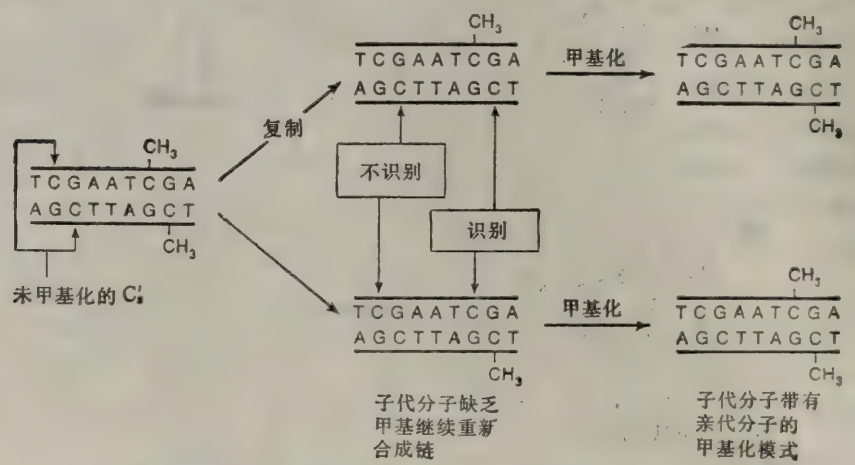


图 23-19 亲代 DNA 的甲基化作用遗传到子代中的模式的图解。只有与已甲基化的序列互补的 CpG 序列中的 C 才能被甲基化。



因为甲基化是脊椎动物中的普遍现象,而且其发生频率因进化程度低下而较低,所以可以假设:甲基化是在进化发生到较高级的阶段才出现的。因而可以期望它只对很近期才进化产生的基因有调控效应。这可以解释为什么甲基化对不同的哺乳动物基因有不同的效果。显然,在得到更多的事实材料之前,甲基化的作用仍将是一个有争议的问题。

未分化的细胞和前体细胞常常要发生 DNA 复制和细胞分裂。如果甲基化的确在某些细胞类型中阻止基因表达,那么在 DNA 复制期间,一个有抑制作用的甲基化位点一定会遗传下去,这种位点在其子链中也必然是甲基化位点。关于这一点的一个可能的机理是,除非甲基化因某种原因被阻止,某些序列永远是甲基化的。然而事实似乎并非如此;而且子链上的甲基化序列与亲代链上的甲基化序列并不相同,而是与亲代链中的甲基化序列互补。某些甲基化酶的一个特点说明了甲基化的亲代链是如何地指导相应的子链序列发生甲基化的:也就是说,这些酶只能使特定序列中的 CpG 甲基化,而且只有在反向链上的 CpG 已经甲基化时,在互补链上才能发在甲基化反应(图 23-19)。这样, DNA 亲代链的甲基化模式才被子链所继承。

有人用高度甲基化来解释 X 染色体的失活。人类男性的性染色体的组成是 XY,女性的性染色体是 XX。如果有一位男性,他有一个以上的 X 染色体(如 XXY),那么他在生理上和发育上都会有很大的障碍,这种人通常智力低下。这提示人们:了解 X 染色体的基因剂量是十分重要的。在女性发育的早期,所有细胞中都有一条在转录上是沉默不动没有转录活性的 X 染色体。推测用这种方式使男性和女性两者的 X 基因剂量维持相等。至于失活哪一条 X 染色体则不是定向选择,而是随机无序选择的结果。关于 X 染色体的失活,有以下证据:(1) 一条 X 染色体高度浓缩并且异染质化;(2) 如果女性的一些基因是杂合子,那么她将表现出与这些镶嵌于 X 的连锁基因的相应性状:一条失去活性的 X 不表现出功能,另一条则表现出基因应有的功能。如果一女性的一个控制汗腺的基因是杂合子,她的皮肤有的区域不能出汗(因为野生型汗腺基因的 X 染色体失去活性,不表现功能);有的区域却能够出汗(因为具有等位突变的汗腺基因的 X 染色体不失去活性,它表现出汗腺基因的活性,也就有功能)。有人提出,失活的 X 染色体是被甲基化了。对是否是甲基化还没有直接的观测方法,主要的证据来自以下实验:如果把一个对许多 X 连锁基因都是杂合子的雌性细胞系列中分离得到的 DNA 用来转化,结果是一个 X 染色体上的基因能够表达,而另一个 X 染色体上的基因不能表达。然而,如果将这一细胞株放在含有 5-氮杂胞嘧啶的培养基中生长几个世代,然后再把从这个培养物中分离得到的 DNA 用于转化,则两个 X 染色体上的基因都是有功能的。因此,通过在 5-氮杂-胞嘧啶中复制,也就是通过对甲基化过程进行抑制,原先失活的 X 染色体又被活化了。事实上,对这一实验的解释曾有过很大争论。

### 互补的转录单位

我们曾多次提到,通常每一 DNA 片段只有一条链被转录。然而,已发现有几个 DNA 片段两条链都被转录。这并不是说两个基因在两条链上完全重叠,而是两个反向转录的基因的起始或终止部分相重叠。这两起反向转录通常在不同程度上互相影响,而这种相互干扰能起调控作用。例如,基因 A 的转录能抑制互补链上基因 B 转录的起始,因

此停止 A 的转录能使互补链上基因 B 的转录开始。在腺病毒中，曾观察到转录终止以及随后的加工处理的作用。在这一事例中两个转录单位在终止区域相重叠。其中一个单位 (X) 的转录能引起另一单位 (Y) 的初级转录片段的过早终止。这种过早终止 (以及随后的多腺苷酸化) 决定了加工的方式，也就决定了以后的拼接方式和由此得出的 mRNA 序列。在这一病毒的生活周期后期，不再形成 X 转录子，因此 Y 的转录也就不被终止；这一更长的初级转录子以另一种方式加工，这种方式使得这一转录单位内的其他基因能够表达。

## 加工过程的调节

转录起始后，最终产生一个初级转录子。在高等真核生物中初级转录子经加工成为一个 mRNA。合成两种肌球蛋白 (称为碱性轻链 LC1 和 LC3) 的鸡骨骼肌细胞就是一个例子。肌球蛋白的基因有两个不同的 TATA 序列，能产生两个不同的初级转录子。这两种转录子以不同方式加工，形成编码不同蛋白质的 mRNA 分子 (见图 23-20)。在果蝇中 mRNA 以四种不同方式加工，具体使用哪种模式取决于果蝇的发育阶段。蛹产生一种肌球蛋白，而晚胚期及幼虫期又产生另一种。加工模式是如何转变的，至今仍是谜。

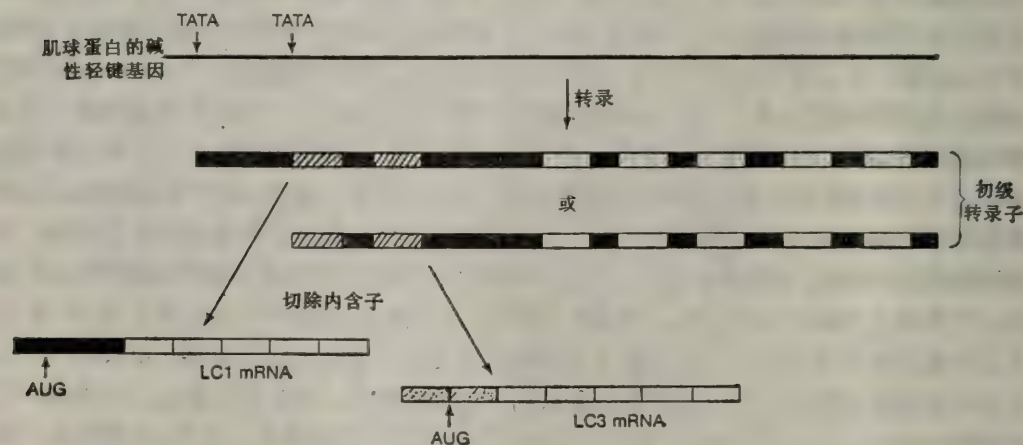


图 23-20 鸡的 LC1/LC3 基因。两个不同的 TATA 序列引起产生不同的初级转录子。这些转录子有着相同的编码序列。删除两种内含子造成形成不同的 mRNA 分子。这些 mRNA 分子编码出具有不同氨基末端而有相同羧基末端的蛋白质分子。

从腺病毒中发现了一些最好的加工调节的例子 (见第二十二章真核病毒)。腺病毒只有很少几个启动子，却产生了大量的初级转录子，并由此产生大量 mRNA (见第二十二章图 22-35)。这样，通过把所有读码框架中的 DNA 都用于翻译，并把不同加工模式所造成的不同 DNA 区域也都用于合成蛋白质，就使很少量的 DNA 得到充分的利用。每种 mRNA 的量则是根据感染后的时间顺序分别受到调节的。例如，E1B 区域的初级转录子以不同方式加工产生几种不同的 mRNA。其中两个见图 23-21。mRNA1 和 mRNA2 都是在初级转录子产生后不久形成的，虽然 mRNA2 的量要大一些。在病毒生活



周期后期, mRNA2 很多,而 mRNA1 的合成停止了。如果在拼接模式转变之前加入蛋白质合成抑制剂环己亚胺,转变就不会发生。这一抑制说明,一种新合成的蛋白质是这个转变的正调节因子 (positive effector), 环己亚胺对于后期其他初级转录子所产生的 mRNA 也有类似效果。

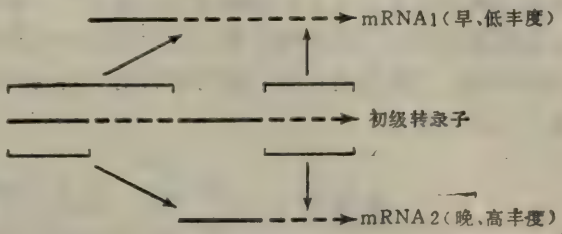


图 23-21 腺病毒 E1B 初级转录子的两种不同的拼接方式。

经常见到这样的情况: 两种细胞合成同一种蛋白质, 虽然蛋白质是由两种不同类型细胞中的同一种基因生成的, 但两者合成出来的蛋白质的量很不相同。这种现象经常伴随着还有其他的 mRNA 分子, 这些不同的 mRNA 分子被翻译的效率也不相同。在大鼠合成  $\alpha$  淀粉酶的过程中, 通过不同的加工序列, 同一基因产生了不同的 mRNA。虽然在动物的不同器官中能转录出相同的编码序列, 但是在鼠的唾液腺中合成出的淀粉酶的量大大地超过肝脏中所合成的。在两种细胞中, 能合成出相同的初级转录子, 但使用了两种不同的拼接加工机理。初级转录子的起始部分见图 23-22。编码序列从外显子 2 中的第 50 个碱基对开始, 通过与外显子 3 及后面的外显子相连接而形成。在唾液腺中, 初级转录子的加工使外显子 S 与外显子 2 相联 (也就是说, 外显子 L 作为内含子 1 和 2 的一部分被切除)。在肝脏中, 因为外显子 S 与内含子 1 和先导序列 L 一起被切除, 所以外显子 L 与外显子 2 相联。外显子 S 和 L 成为淀粉酶的两种先导序列, 这样就产生出不同的翻译速度。

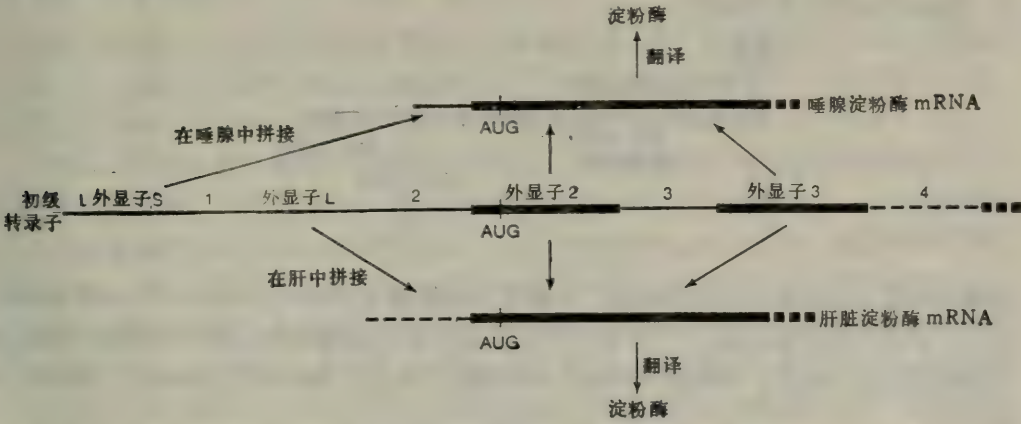


图 23-22 在小鼠唾液腺和肝脏细胞中通过不同的剪接方式产生不同的淀粉酶 mRNA。先导 L 区及内含子用细黑线表示。外显子用粗黑线表示。编码序列于外显子 2 的 AUG 密码子处开始。

另一种类型的加工调节涉及到选择不同的多腺苷酸化位点。不同组织中激素钙调素的不等速合成就是一个例子。另一个例子是两种免疫球蛋白  $\mu$  链的合成。在合成两种  $\mu$  链的过程中, 同样的初级转录子的不同加工过程都包含有根本不同的内含子切除模式 (即拼接模式)。但这必然需要先从初级转录子上选择不同的 poly A 位点。也就是说, 如果选择更加靠近启动子的 poly A 位点时, 那么就失去了在较大的初级转录子中能使用的一个剪接位点。于是产生了另一种拼接模式。因此, 在这两种情况下可能翻译出略有不同

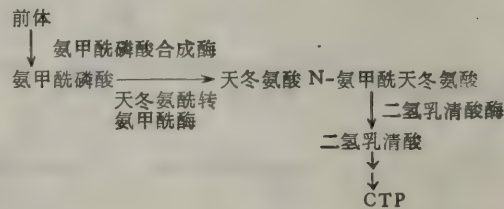
同的蛋白质。

腺病毒生活周期后期的所有 mRNA 都只有一个启动子。初级转录子在五个主要的多腺苷酸化位点上终止(图 22-35)。每一个终止位点通过使某些特殊的内含子或内含子末端的存在或丢失来影响拼接。五个位点的使用频率不同,从而使编码各基因的 mRNA 的量不同。这是在病毒的生活周期后期决定病毒的不同结构蛋白被合成出不同的相对量的主要机理。

### 多聚蛋白质: 同一 mRNA 产生多种蛋白质

在原核细胞中,对几个基因产物的协同调节是通过控制对编码这些蛋白质一个多顺反子 mRNA 分子的合成来实现的。在真核细胞中,与此类似的机理是多聚蛋白质(polyprotein)的合成。多聚蛋白质是一个很大的多肽,在翻译后能被切割生成好几种蛋白质。每一种蛋白质可以被看作是一个基因的产物。在这个体系中,形成多聚蛋白质的 DNA 中各个基因的编码序列不是被终止密码和起始密码隔开,而是被专一性的氨基酸序列隔开,这些序列能被专一性的蛋白质切割酶作为切割位点识别出来。已见到的多聚蛋白质最多有 8 个切点(脊髓灰质炎病毒,见第二十一章图 21-12);在各个切割位点上并不是同时切割,而是以特定的顺序进行切割。这种机理能够保证各组成蛋白质等量地合成,就象以前在非洲爪蟾 rRNA 的形成中所见到的。而且,某些位点切割的延迟能够为各个蛋白质的合成确定时间顺序,这是动物病毒常用的机理(第二十章图 22-22)。

已知有很多多聚蛋白质的例子。在 RNA 前体 UTP 和 CTP 的生物合成途径中的某一阶段有以下的反应顺序:



在细菌中以上三种酶是分别合成的。在酵母和链孢霉(*Neurospora*)中只合成两种蛋白质:一种是二氢乳清酸酶;另一种是一个大的蛋白质,它被切割成氨甲酰磷酸合成酶和天冬氨酸转氨甲酰酶。哺乳动物则合成一种包括三部分的蛋白质,它随后被切割成三种酶。

有一些多聚蛋白质在不同组织中以不同方式受到切割。前促黑色素促肾上腺皮质激素原(proopiomelanocortin)就是一个例子。这个多聚蛋白质是脑下垂体中合成的多种激素的来源。在丘脑后叶,这个多聚蛋白质先从C末端切割释放出 $\beta$ -促脂激素,见图 23-22。然后切割N端,释放促肾上腺皮质激素(ACTH)。在这一组织中不发生其他切割。然而在中叶中 $\beta$ -促脂激素进一步切割,释放C末端的多肽: $\beta$ -内啡肽。ACTH也被切割,产生 $\alpha$ -促黑激素。总之,在后叶中一次切割产生两种激素;在中叶中第二次切割把上述两种激素再转变为另两种新的激素。

以上所述的各种多聚蛋白质都被切割成等量的几种不同蛋白质(每种蛋白质一个拷



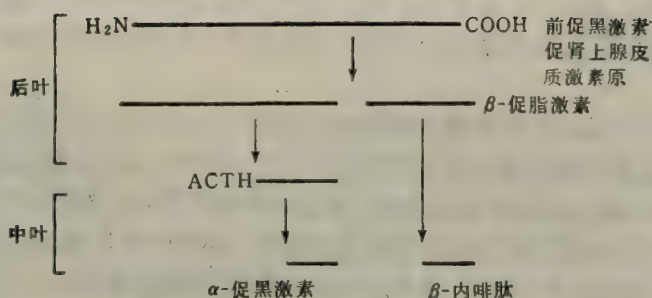


图 23-23 切割垂体的多聚蛋白质前促黑激素促肾上腺皮质激素原产生四个激素。β-促脂激素及 ACTH 在垂体后叶中形成，α-促黑激素和 β-内啡肽在垂体中叶中产生。垂直的箭头指出切割位点。

贝)。然而并非所有的多聚蛋白质都是这样的。例如，内啡肽的前体就包含着六个 Met-内啡肽和一个 Leu-内啡肽。

## 翻译过程的调控

翻译过程的控制，是指调节控制成熟了的 mRNA 分子的翻译次数。mRNA 分子的翻译可以通过以下三种方式来调节：(1) mRNA 的寿命；(2) 起始翻译的概率；(3) 控制蛋白质合成的速度。在这一部分中，我们接着讨论这三种调控模式的例子。

### 延长 mRNA 的寿命

丝蚕 (*Bombyx mori*) 的丝腺主要合成一种蛋白质，即丝纤维蛋白。丝蚕筑造蚕茧达数天之久，因此丝纤维蛋白的合成量必须很大，而对合成速度却没有什麼要求。因此，丝蚕可以利用寿命很长的纤维蛋白 mRNA 分子去制造丝纤维蛋白。

一般认为丝纤维蛋白基因有一个很强的启动子，当蚕接收到一个翻译的起始信号(其具体性质尚不明)后，在几天内就能合成出大约  $10^4$  个丝纤维蛋白 mRNA。一般真核细胞 mRNA 在降解前的平均寿命大约有 3 小时。然而，丝纤维蛋白的 mRNA 却能存在几天，在这几天内，mRNA 分子可以不断地翻译，每一个 mRNA 分子能产生  $10^5$  个丝纤维蛋白分子。这样，在四天内每个基因就可以合成出  $10^9$  个蛋白质分子。如果产生丝纤维蛋白的细胞是双倍体，那么每一套双倍体染色质就能合成出  $2 \times 10^9$  个蛋白质分子。整个丝腺若在四天中应合成 300 微克或  $10^{15}$  个分子的丝纤维蛋白，那么大约共需要  $5 \times 10^5$  个双倍体细胞。然而，丝腺在发育过程中有一独特的性质，即丝腺是从一个双倍体细胞发育而来的。当原初细胞的染色体加倍并通过有丝分裂而分开时，细胞并不分裂，这样就产生四倍体细胞。以后染色体继续地加倍和有丝分裂一直继续地进行，但总是没有细胞分裂，直到丝腺成熟为止。这时丝腺已是一个大约有  $10^6$  倍染色体的大细胞。这  $10^6$  套染色体中的每一个染色体都各有一个有活性的丝纤维蛋白基因，此基因能合成  $10^9$  个丝纤维蛋白分子。因此，这个与众不同的细胞就能合成  $10^6 \times 10^9 = 10^{15}$  个蛋白质分子。目前还不知道在形成丝腺的过程中细胞分裂为什么会受到抑制。

在高度分化的细胞中，通过延长 mRNA 的寿命来合成大量的某种单一类型的蛋白

质是一种普遍现象。例如,在合成卵白蛋白(以便形成蛋白)的鸡的卵巢细胞内,虽然在每一套单倍体染色体中只具有单拷贝的卵白蛋白基因,但是细胞质 mRNA 却是长寿物的。

在这里有必要把这种有时被称为翻译放大过程(translational amplification)的蛋白质合成方式与非洲爪蟾卵母合成所需要的  $10^{12}$  个 rRNA 分子的基因扩增过程做一个比较。丝蚕之所以能合成大量蛋白质分子是因为有两步扩增,即以 DNA 为模板合成  $10^4$  个 mRNA 分子从而使之扩增  $10^4$  倍,然后每个 mRNA 再反复地翻译,从而又扩增  $10^5$  倍,共  $10^9$  倍。然而,在 rRNA 转录的合成过程中,转录是最后的一步,没有翻译的过程。因此,为了得到一个大的总扩增系数,在转录之前必然有一个扩增步骤,也就是说,基因本身必然被大大地扩增。我们已经提到,丝蚕合成丝纤维蛋白并不需要速度很快。然而,在其他生物或其他细胞系统中,有时需要快速地合成蛋白质。这时,仅靠翻译扩增就不够了。例如,在受精卵分裂时,必须以每分钟约  $3 \times 10^5$  分子的速度去合成构成新的染色质所需的组蛋白,才能跟上 DNA 合成的速度。如果按照通常的合成速度,每分钟大约只能合成 1 000 个分子。受精卵中蛋白质合成之所以达到了所要求的高速度,是由于每一个组蛋白基因都产生了几百个初始拷贝。请注意这与本章前面讨论过的基因扩增不同,因为在所有的细胞中这个基因的拷贝数都是这样的。

各种专一性的 mRNA 分子有着不同的半寿期,这一点在腺病毒的生活周期中很重要(第二十一章)。转录单位 1B 能被转录,产生两种 mRNA 分子: 分别是 14S 和 20S。这种区别是由于初级转录子有不同的剪接方式所造成的。在病毒生活周期的后期,14S mRNA 的量与 20S mRNA 的量的比例远较前期的要大,这是因为后期 14S mRNA 的半寿期变长的结果。

有时转录的调控与翻译的调控是相结合的。例如,在哺乳动物组织中,牛奶中的蛋白质合成要求胰岛素(能调节很多物质的合成)和催乳素(另一种激素)共同参与,以促进酪蛋白(牛奶蛋白质)的合成。这两种激素,对于起始酪蛋白基因的转录都是必要的;此外,催乳素还能增加酪蛋白 mRNA 的寿命。但是对于这些作用的机理目前还不清楚。

### 控制翻译的起始

翻译调控的另一个例子是被掩盖着的 mRNA (masked mRNA) 的调控。未受精的海胆卵能以核蛋白颗粒的形式很容易地储存大量 mRNA 达几个月之久,这些 RNA 是无活性的,但是在受精之后几分钟之内就可以开始翻译这些 mRNA 分子。目前对于如何稳定住 mRNA 以保护 mRNA 不受核酸酶作用,以及如何使这些 mRNA 活化的机理尚不清楚。

因为未受精的成熟的卵细胞只需要保持自己的生存,不需要生长或改变状态,所以卵中的蛋白质合成速度一般来说是很低的。这并不是因为 mRNA 供应不足,而是因为起始翻译所需的某些因素受到限制。这个结论是从一个特殊的实验中推出来的,其数据见图 23-24。蟾蜍卵很大,用皮下注射器就能很容易地把小体积的液体注入。如果把珠蛋白的 mRNA(从红细胞中分离而得)与一种放射性氨基酸一起注入蟾蜍卵,就能够合成出放射性球蛋白。图 23-24(a)表示逐渐增大注入的珠蛋白 mRNA 的量所得的结果。珠蛋白的 mRNA 加得越多,所合成的放射性珠蛋白也就越多,但正常的卵蛋白的合成量却减少了。这表



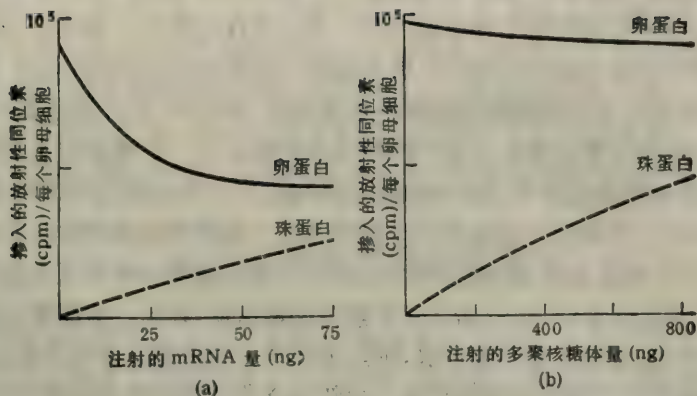


图 23-24 注射网织红细胞珠蛋白 mRNA (a 行)或网织红细胞多聚核糖体 (b 行)对放射性氨基酸掺入到卵母细胞蛋白质的影响。

卵母细胞利用 mRNA 的能力已达到饱和。再加入外源 mRNA 就会与内源 mRNA 竞争蛋白质合成的装置。然而,如果注入的是含有珠蛋白 mRNA 的网织红细胞的多聚核糖体,而不是纯化的珠蛋白 mRNA,那么卵自身蛋白的合成减少并不很显著[图 23-24 (b)]。这就是说氨基酸、延长因子以及终止因子的供应很充足。综合这两个实验,可以看出,限速因子或者是多聚核糖体的数量,或者是形成多聚核糖体所必须的某种分子。而核糖体在这里是过剩的,所以起作用的很可能是某种其他因子而不是核糖体本身。这种因子被称为补充因子 (recruitment factor)。

某些蛋白质的合成是由 mRNA 上的蛋白质的直接作用来调控的。例如,一种免疫球蛋白通过自我阻遏翻译来使其浓度保持恒定。这种蛋白质与其他所有的免疫球蛋白一样,都有两个 H 链和两个 L 链。这个四体分子专一性地与 H 链的 mRNA 结合,从而阻遏翻译的起始。关于 L 链合成是如何控制的,目前还不清楚。

## 自我调节

在第十四章我们看到原核细胞中有许多基因,尤其是那些不断表达的基因进行自我调节 (autoregulation)。在真核细胞中也发现少量自我调节的例子。真核组织中大部分细胞具有由微管蛋白 (tubulin) 构成的称之为微管纤维网络的结构。微管蛋白是纺锤体的主要成分。纺锤体是在有丝分裂中使染色体向细胞两端分开的细胞器。秋水仙碱 (colchicine) 和长春花碱 (vinblastine) 两种药物能抑制微管蛋白的多聚化作用,从而增加细胞内游离微管蛋白的浓度。向组织培养的细胞中加入这两种药物中的任一种,都能使微管蛋白 mRNA 消失。而且,如果用微量注射针头把微管蛋白注入到哺乳动物细胞中,新的微管蛋白的合成速度就大大减慢。这些实验说明微管蛋白能抑制其自身的合成。然而,在以上两个实验中微管蛋白的初级转录子的合成都未受到阻碍,因此,抑制或是发生在 RNA 拼接水平上,或是发生在翻译水平上。目前还没有能确定这方面的进一步的信息。

## 蛋白质合成总速度的调节

只合成一种蛋白质的细胞,有时通过控制蛋白质合成的总速度来调节这种蛋白质的

合成。网织红细胞合成血红蛋白就是通过这种方法进行调节的。在分化出红细胞的分化过程中，网织红细胞是倒数第二种细胞型态。在大多数生物中，网织红细胞都没有细胞核，细胞核在原红细胞（erythroblast）转变成网织红细胞（reticulocyte）时被排出细胞了。网织红细胞依靠原红细胞的细胞核所转录出的 mRNA 分子来合成  $\alpha$ -和  $\beta$ -珠蛋白亚基。另一个酶系统负责合成血红蛋白的辅基血红素。每一个珠蛋白亚基中有一个血红素，而且为了要得到最佳的效率，每一种类型的珠蛋白（ $\alpha$  和  $\beta$ ）的合成速度都应该是血红素合成速度的两倍。正确的合成速度的比例是通过一种两相的调控体系来维持的。在这个调控体系中：（1）血红素阻遏其自身的合成；（2）血红素使一种珠蛋白合成的抑制物失去活性。在这个调控体系作用下，如果血红素比珠蛋白多，则停止血红素的合成，促进珠蛋白的合成；反之，如果血红素的量不够，则阻遏珠蛋白的合成，并再开始血红素的合成。这样，血红素与珠蛋白的量始终保持平衡。血红素合成的控制机理目前还不清楚，但是如上所述，对于血红素是如何控制珠蛋白的合成我们已经有了较好的了解，下面将要介绍。

兔网织红细胞没有核，因而也就没有 DNA。所以这些细胞中的珠蛋白的合成无法在转录水平上受到调控，而只能在翻译水平上被调节。调节模式如图 23-25 所示。要起始翻译过程必须先形成一个三聚体复合物，它包括  $tRNA_{init}$ 、起始因子 eIF-2 和一个被称为促真核起始因子蛋白的分子，即 ESP（eIF-2-stimulating protein）。网织红细胞中有一个被称为控制血红素阻遏物即 HCR（hemin-controlled repressor）的蛋白激酶（它是一种蛋白质磷酸化酶）。这个阻遏物能够使 eIF-2 的小亚基磷酸化，从而阻止 eIF-2 与 ESP 形成复合物。而 HCR 本身也受一个依赖于 cAMP 的蛋白激酶激活。这个激酶有四个亚基，两个相同的亚基称为 R，另两个称为 C，所以此酶可以用  $R_2C_2$  来表示。R 与 C 亚基分别具有调控和催化作用。两个 R 亚基都能与 cAMP 结合。在有 cAMP 时， $R_2C_2$  解离，释放出 C 亚基，这样的 C 亚基能够使 HCR 磷酸化。然后被激活的 HCR 使 eIF-2 磷酸化，从而阻止蛋白质合成的起始。血红素能与 R 亚基结合，使之发生构象变化，从而阻碍 R 与 cAMP 结合。因此，如果血红素过剩，HCR 就不能被激活，蛋白质

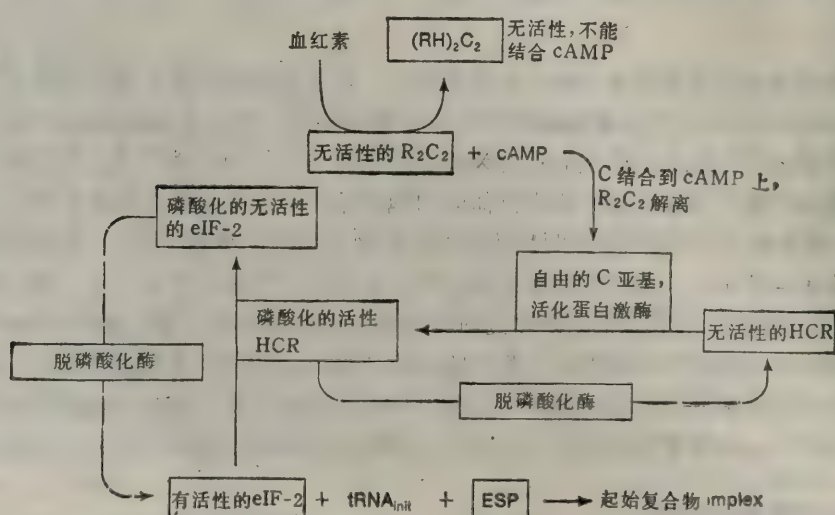


图 23-25 氧高铁血红素对网织红细胞中蛋白质合成的调节。



合成就能够起始。

值得一提的是,珠蛋白和 HCR 都能与血红素结合,只是珠蛋白的结合更强一些。因此,当珠蛋白合成出来后,立即与血红素结合,从而减少游离血红素的浓度,使 HCR 能被活化。接着珠蛋白的合成就受到抑制,直到合成出更多的血红素,珠蛋白的合成才得到恢复。请注意,这里并不是专一性地控制珠蛋白合成的开与关,而是对整个蛋白质合成的速度进行调控。这样的调控作用只有在仅仅合成一种蛋白质的细胞中才能做到。

## 光对植物基因表达的调节

许多高等植物在黑暗中生长几天后,细胞会失去叶绿体,也就失去了绿色。这种丢失(称为白化)是由于失去了催化叶绿体合成的酶,而新酶又不能再合成所造成的。当白化的植物重新光照后,几个小时就能开始合成 60 种以上光合作用所需的酶、叶绿体 rRNA 和叶绿体。已经仔细地研究过双磷酸核酮糖羧化酶即 RDC (ribulose diphosphate carboxylase)——一种光合作用酶的一个亚基的光照再合成过程。这个亚基的合成是由一种称为光敏色素 (phytochrome) 的分子为中介的,光敏色素这个蛋白质常有一个共价结合的光吸收色素。在黑暗中光敏色素是没有活性的;在光照下它就转变成为有活性的形式。有活性的光敏色素能够参与转录编码 RDC 小亚基基因(以及一些其他基因)过程的诱导作用。其作用机理还不知道,但通常认为这种蛋白质是一个光激活的转录因子。大亚基的合成也由光敏色素调节,一些其他蛋白质也起作用,但是对这些则了解甚少。

光敏色素同时也调节其自身的合成(自我调节)。在黑暗中,光敏色素的基因能被转录,但当这个蛋白质被光激活之后,其转录受到抑制。注意,光在这里的作用与光对光合酶类的光诱导激活作用正好是相反的。

集群藻类 (volvox) 的基因表达也是由光调节的。在有光照和黑暗时,细胞中的 mRNA 量基本上是一样的。但是由黑暗向光照的转变能够引起翻译的大幅度提高。关于这个现象以及藻类中类似的其他一些现象的机理目前还是一个谜。

## 卵黄蛋白原合成调节的一个例子

卵黄蛋白原 (vitellogenin) 是卵黄中的卵黄高磷蛋白 (phosvitin) 和卵黄脂磷蛋白 (lipovitellin) 两种蛋白质的前体。卵黄蛋白原的特别有趣之处,在于其合成过程在几处受到调控,而且可以在这种蛋白质的合成途径中见到前面已讨论过的许多调控特点。研究得最详尽的是非洲爪蟾 (*Xenopus laevis*) 中的卵黄蛋白原的合成。

在卵黄蛋白原基因家族中至少有四个基因,分别称为 A1、A2、B1 和 B2。A 类基因的翻译产物与 B 类基因的翻译产物差别较大,约 20% 的氨基酸有差异。而亚组 1 基因的产物与亚组 2 基因的产物则只有约 5% 的氨基酸有差异。四种基因产物的合成受到同一个体系的调节。然而,多聚蛋白质基因产物的切割模式却是不相同的。

### 卵黄蛋白的合成

卵黄高磷蛋白和卵黄脂磷蛋白的合成途径见图 23-26。合成的三个主要阶段如下:

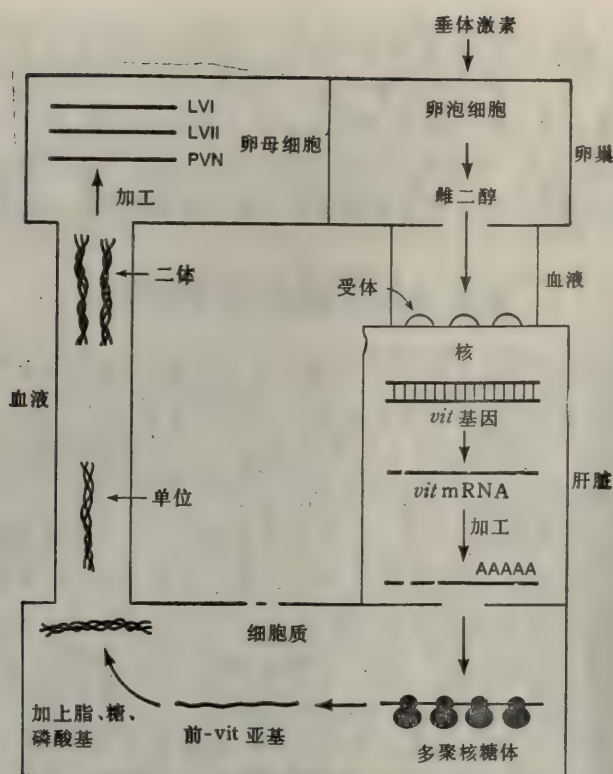


图 23-26 雌二醇对合成卵黄蛋白原的促进及其再加工形成卵黄高磷蛋白原 (PVN)、卵黄脂磷蛋白 I 及 II (LV) 和卵黄磷蛋白 (PVT) 途径的概要图。图中指出了在卵母细胞中 A 类型的卵黄高磷蛋白原的加工过程。B 类型则有四种蛋白质,即 LVI、LVII、PVTI 及 PVTII。

(1) 卵黄蛋白在成熟的雌性动物肝中合成。合成过程受雌性性激素促进。(2) 合成的卵黄蛋白分泌到血液中,运送到卵巢。(3) 在卵巢中,发育中的卵(卵母细胞)选择性地卵黄蛋白原吸收,然后再切割成为卵黄蛋白。

卵黄蛋白原的合成细节如下:

卵黄蛋白的合成是由控制转录的激素调节的。转录子经过加工后被翻译出一个多聚蛋白质,进一步再受到修饰。最后从修饰过的多聚蛋白质上切割下一个多肽链。其步骤如下(图 23-26):

1. 卵黄蛋白原合成的最初刺激来自下丘脑(脑的一部分)。这个刺激使脑下垂体分泌促性腺激素。

2. 促性腺激素被发育中的卵外面包裹着的卵泡吸收。卵泡细胞作出的反应是分泌雌激素到血液中。

3. 雌激素结合到肝细胞膜中的受体上面,肝细胞作出反应,合成卵黄蛋白原。雌激素的这个作用使得卵黄蛋白原的合成成为一个很有价值的实验体系。这是因为母鸡与公鸡的肝细胞都能合成卵黄蛋白原,但是只有母鸡才能合成雌激素。如果使用公鸡作实验,就有可能通过往血液中注射雌激素而随时使这个体系开始合成卵黄蛋白原,从而使人们能



够有控制地、仔细地观察与研究这个蛋白质合成的初期过程。

4. 雌激素被运送到肝细胞的核中,在细胞核中开动卵黄蛋白原基因的转录。雌激素对卵黄蛋白原合成的激发模式与以前所叙述的激素激发转录基本机理有一些区别。首次接触雌激素时,每个肝细胞的核上大约有 200 个雌激素受体。12 个小时以后,由于又新合成了许多受体,受体数增加到 1000 个,细胞保持这个受体数目达 8 天之久。受体数目的增加使人们可以解释下面的一现象:如果在头一次给一个动物注射小剂量的雌激素,几天以后不再合成卵黄蛋白原时,再第二次注射雌激素,于是发现卵黄蛋白原 mRNA 的合成速度比第一次大多了。

5. 转录子经加帽,加 poly A 尾,再由加工酶除去内含子形成 mRNA, mRNA 被运送到细胞质中之后,形成多聚核糖体。

令人感兴趣的是有了雌激素,卵黄蛋白原 mRNA 的寿命得以延长。只要有雌激素存在, mRNA 的半寿期大约是 24 小时。但是如果用雌激素连续处理几天之后再撤去雌激素, mRNA 的半寿期缩短到 3 小时。

6. 加工过的 mRNA 开始被翻译。然而,虽然整体的蛋白质合成是在继续正常地进行,但是翻译过程并不立刻开始。据认为,需要有一个或几个翻译因子才能起始卵黄蛋白原 mRNA 的翻译过程,而这些因子也是在雌激素的作用下才合成出来的。关于这一点的证据如下:如果对雄性动物给予雌激素,卵黄蛋白原开始合成之后,撤去雌激素,卵黄蛋白原的合成也就停止了。这是因为不再继续为卵黄蛋白原合成其新的 mRNA;而已有的 mRNA 又已经降解。如果此时再给予雌激素,则当卵黄蛋白原的 mRNA 刚被合成出来,就立刻被翻译,此处不存在一个延迟阶段。这是因为已经具有翻译因子了。

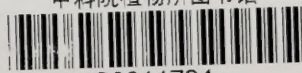
7. 翻译能力加强。一旦卵黄蛋白原的合成开始以后,注射雌激素使肝细胞能够比以前更快地合成全部蛋白质。这是通过一个实验证实的。在这个实验中,对注射雌激素前、后的两种动物中所得到的两份肝细胞提取液进行测定,比较了它们受 polyU 中介而合成多聚苯丙氨酸的能力。人们发现,从经过雌激素处理的雄性动物中提取的肝细胞提取液,合成多聚苯丙氨酸的能力高得多。因此认为翻译受到两种效应使肝细胞能够合成出大量的卵黄蛋白原:(1)总蛋白质合成的速度加快,(2)某些细胞质因子使卵黄蛋白原 mRNA 能够很有效地与其他 mRNA 竞争,从而使卵黄蛋白原成为这一类蛋白质合成中的主要产物。

8. 肝细胞中发生的最后一步,是对蛋白质进行修饰:加磷酸基团,加脂基,加糖基。

9. 已完成合成过程的卵黄蛋白原单体被分泌到血液中。在那里,这种修饰过的蛋白质分子又聚合成双体,这就是通常所指的卵黄蛋白原的分子。

10. 卵母细胞从血液中吸收这种双体分子。这种双体分子是一种多聚蛋白质,它在卵母细胞中又被切割。A 类和 B 类蛋白的切割方式不同。A 类多聚蛋白质被切割成卵黄高磷蛋白(分子量  $M = 35\ 000$ )、卵黄脂磷蛋白 I ( $M = 115\ 000$ ) 和卵黄脂磷蛋白 II ( $M = 31\ 000$ ); B 类多聚蛋白质则形成卵黄脂磷蛋白 I 和 II, 以及 phosvette I ( $M = 14\ 000$ ) 和 phosvette II ( $M = 19\ 000$ )。存在多型卵黄蛋白原的重要意义目前还不清楚,但是在受到切割之后,以上两种情况下产生的蛋白质是被装配在卵黄小板(yolk platlet)中,以形成成熟的卵。





## 卵清白蛋白的协同合成

鸟类和两栖类的卵有卵黄和卵清。卵黄与卵白的蛋白质合成应该是协同调控的,因为这两者唇齿相依,事实正是如此,调节因子仍是雌激素。雌激素不仅起始卵黄蛋白原的合成,也起始卵白蛋白(本章前面已提到过,这是卵清蛋白中主要的蛋白质)的合成。

## 参 考 文 献

- Axel, R., et al. 1979. *Eukaryotic Gene Regulation*. Academic Press.
- Bienz, M. 1985. "Transient and developmental activation of heat-shock genes." *Trends Biochem. Sci.*, 10, 157.
- Bostock, C. 1980. "A function for satellite DNA?" *Trends Biochem. Sci.*, 5, 117.
- Brent, R. 1985. "Repression of transcription in yeast." *Cell*, 42, 3.
- Brown, D. 1981. "Gene expression in eukaryotes." *Science*, 211, 667.
- Capasso, O., and N. Heintz. 1985. "Regulated expression of mammalian histone H4 genes *in vivo* requires a trans-acting transcription factor." *Proc. Nat. Acad. Sci.*, 82, 5622.
- Chisholm, R. 1983. "Gene amplification during development." In *DNA Makes RNA Makes Protein*. (T. Hunt et al., eds.) Elsevier Biomedical.
- Collins, F. S., and S. M. Weissman. 1984. "The molecular genetics of human hemoglobin." *Prog. Nucl. Acid. Res. Mol. Biol.*, 31, 317.
- de Crombrughe, B., and I. Pastan. 1983. "Structure and regulation of a collagen gene." In *DNA Makes RNA Makes Protein*. (T. Hunt et al., eds.) Elsevier Biomedical.
- Dynan, W. S., and R. Tjian. 1985. "Control of eukaryotic mRNA synthesis by sequence-specific DNA-binding proteins." *Nature*, 316, 774.
- Gough, N. 1981. "The rearrangement of immunoglobulin genes." *Trends Biochem. Sci.*, 6, 203.
- Guarente, L. 1984. "Yeast promoters: positive and negative elements." *Cell*, 36, 799.
- Hood, L. E., I. L. Weissman, and W. B. Wood. 1978. *Immunology*. Benjamin.
- Jelinek, W. R., and C. W. Schmid. 1982. "Repetitive sequences in eukaryotic DNA and their expression." *Ann. Rev. Biochem.*, 51, 813.
- Karlsson, S., and A. W. Nienhaus. 1985. "Developmental regulation of human globin genes." *Ann. Rev. Biochem.*, 54, 1071.
- Kolata, G. 1981. "Gene regulation through chromosome structure." *Science*, 214, 775.
- Kolata, G. 1984. "New clues to gene regulation." *Science*, 224, 588.
- Kolata, G. 1985. "Fitting methylation into development." *Science*, 228, 1183.
- Leighton, T., and W. F. Loomis, eds. 1982. *The Molecular Genetics of Development*. Academic Press.
- Long, E. O., and I. B. Dawid. 1980. "Repeated genes in eukaryotes." *Ann. Rev. Biochem.*, 49, 727.
- Marx, J. L. 1981. "Antibodies: getting their genes together." *Science*, 212, 1015.
- Numa, S., and S. Nakanishi. 1983. "Corticotropin- $\beta$ -lipotropin precursor." In *DNA Makes RNA Makes Protein*. (T. Hunt et al., eds.) Elsevier Biomedical.
- Ochoa, S. and C. de Haro. 1979. "Regulation of protein synthesis in eukaryotes." *Ann. Rev. Biochem.*, 48, 549.
- O'Malley, B. W., and W. T. Schroeder. 1976. "The receptors of steroid hormones." *Scient. Amer.*, February, p. 32.
- Revel, M., and Y. Goner. 1978. "Posttranscriptional and translational controls of gene expression in eukaryotes." *Ann. Rev. Biochem.*, 47, 1079.
- Safer, B., and W. F. Anderson. 1978. "The molecular mechanism of hemin synthesis and its regulation in reticulocytes." *CRC Rev. Biochem.*, 6, 261.
- Sakano, H., et al. 1979. "Sequences in the somatic recombination of immunoglobulin light chain genes." *Nature*, 280, 288.
- Seidman, J. G., and P. Leder. 1978. "The arrangement and rearrangement of antibody genes." *Nature*, 276, 790.
- Tata, J. R. 1976. "The expression of the vitellogenin gene." *Cell*, 9, 1.
- Taylor, J. H. 1984. *Methylation and Cellular Differentiation*. Springer-Verlag.
- Wahli, W., et al. 1981. "Vitellogenesis and the vitellogenin gene family." *Science*, 212, 298.
- Weissbrod, S., and H. Weintraub. 1979. "Isolation of a subclass of nuclear proteins responsible for conferring a DNase I-sensitive structure in globin chromatin." *Proc. Nat. Acad. Sci.*, 76, 630.



来源	91.10.26
书价	西单书
单据号	20.90
开票日期	160624
	91.10.25

25752		58.17	
		165	
分子生物学		1991	
实验生物和细胞生物总论 下册			
借者单位	借者姓名	借出日期	还书日期
林学	2991.1.82	97/10/1	于29/6
	93.4.3		
	26		

58.17  
165

### 注 意

- 1 借书到期请即送还。
- 2 请勿在书上批改圈点，折角。
- 3 借去图书如有污损遗失等情形须照章赔偿。

25752

京卡0701

封面设计：丁钟颖

ISBN 7-03-002138-2

Q · 294

定 价： 20.90元

科技新书目：239-043